The general response style from a cross-cultural perspective

He, J.

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

THE GENERAL RESPONSE STYLE FROM A CROSS-CULTURAL PERSPECTIVE

JIA HE

THE GENERAL RESPONSE STYLE FROM A CROSS-CULTURAL PERSPECTIVE

Proefschrift

ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,

in het openbaar te verdedigen
ten overstaan van een door het college
voor promoties aangewezen commissie

in de aula van de Universiteit
op woensdag 4 februari 2015 om 14:15 uur

door

Jia He
geboren op 24 november 1984
te Sichuan, China

PROMOTIECOMMISSIE:

PROMOTOR:	Prof. dr. A. J. R. van de Vijver
COPROMOTOR:	Dr. A. del Carmen Domínguez Espinosa
OVERIGE LEDEN:	Prof. dr. P. B. Smith Prof. dr. Y. H. Poortinga Prof. dr. H. van Herk Dr. C. H. van Wijk Dr. B. Weijters

Table of Content

Chapter 1

Introduction

Response styles, defined as the systematic tendency to respond to questionnaires on some basis other than the target construct, have been studied since the 1950s (Cronbach, 1942, 1950). However, the psychological meaning of response styles and their implications on the validity of data are still under debate, especially in cross-cultural contexts (e.g., Baumgartner & Steenkamp, 2001; Messick, 1991). There are two divergent interpretations of response styles. The first is the traditional and still dominant perspective in which response styles are treated as systematic measurement errors that should be avoided and eliminated as much as possible. An alternative interpretation holds that response styles are a basic way of communicating about oneself, such as the tendency to amplify responses among Latin Americans and to moderate responses among East Asians, so response styles are embedded in the values and personality of respondents and their cultures (P. B. Smith, 2004, 2011). Sixty years of research on response styles have produced ample empirical data, yet conceptual progress is less impressive. The main challenges in response style research, among others, are the different operationalizations of specific response styles, lack of validity measures that are less susceptible to the influence of response styles, and the inconsistency in response style correction effects. This dissertation aims to advance our understanding of response styles from a cross-cultural perspective by (1) integrating different response styles to a general factor, (2) establishing the nomological network of response styles with various validity measures at both individual and cultural level, and (3) exploring the implications of response style effects on the validity of scores in cross-cultural surveys. This chapter first reviews literature on the integration of response styles, their nomological network, effects of response styles on responses, and their corrections in cross-cultural contexts. Then it states the research questions and the outline of this dissertation.

Interrelatedness of Specific Response Styles

The most frequently studied response styles include acquiescent response style (ARS), extreme response style (ERS), midpoint response style (MRS), and socially desirable responding (SDR). ARS is defined as the tendency to agree rather than disagree to propositions in general; ERS is conceptualized as the tendency to endorse the most extreme response categories regardless of content; MRS refers to the tendency to frequently use the midpoint of a scale; and SDR is the tendency to answer questions in a way that makes oneself look good (Paulhus, 1991). ARS, ERS, and MRS are mostly measured indirectly with items of other substantive constructs. Many SDR scales have been developed (Paulhus,

2002), and subdimensions, such as enhancement and denial (Ramanaiah, Schill, & Leung, 1977) and impression management and self-deception (Paulhus, 1984), have been proposed in the literature.

The definitions of these response styles and their correlations with other psychological variables suggest that they are interrelated. ERS, a tendency to be unequivocal with a self-promotion focus, can be viewed as the opposite of MRS, a tendency to be evasive with a prevention focus (Cabooter, 2010). Smith and Fischer (2008) found that ARS was more often endorsed by people with collectivistic values and ERS more by people with individualistic values. Therefore, a negative association between these two can be expected. SDR and ERS have in common that they represent desirable traits related to extroversion and conscientiousness (Austin, Deary, & Egan, 2006; Musek, 2007). Although conceptually related, these four response styles are seldom studied simultaneously. In this dissertation, these four response styles are integrated in a General Response Style (GRS) with ERS and SDR as positive indicators and ARS and MRS as negative indicators at both individual and country level.

Nomological Network of Response Styles

At both individual and cultural level, response styles are found to be associated with various psychological measures, notably personality. At individual level, ARS was associated with impulsiveness and extraversion (Couch & Keniston, 1960). ERS was positively related to intolerance of ambiguity, simplistic thinking, decisiveness, extroversion, and conscientiousness (Austin et al., 2006; Naemi, Beal, & Payne, 2009). MRS was associated with evasiveness (Ayidiya & McClendon, 1990). SDR was associated with the general personality factor which consists of the combination of agreeableness, conscientiousness, extroversion, openness, and emotional stability (Bäckström, 2007; Schermer & MacDougall, 2013). Besides personality, these response styles at individual level are associated with values, emotion regulation, and positive life outcomes (e.g., Bachman & O'Malley, 1984; Lalwani, Shrum, & Chiu, 2009).

At country level, McCrae et al. (2005a) reported a negative association of ARS with conscientiousness. Smith (2011) found a negative association of ARS with openness. Harzing (2006) reported a positive association of extroversion with ERS and a negative one with MRS. Moreover, previous studies found that response styles were related to a cluster of collectivistic values, including collectivism, embeddedness, and traditionalism (versus

secularism) (e.g., Harzing, 2006; P. B. Smith, 2004; van Dijk, Datema, Piggen, Welten, & van de Vijver, 2009).

Two issues have arisen from the search of the nomological network of response styles. Firstly, the validity measures reviewed above are typically from Likert scales, which themselves are likely to be tainted by the same response styles as are being investigated (Bentler, Jackson, & Messick, 1971). Therefore, validity measures that are robust to the effects of response styles are needed to shed light on the nature of these response styles. This dissertation makes use of "hard" cultural indicators such as country affluence, religious denomination, and educational achievement, and a forced-choice format personality measure to validate response styles. Secondly, there seems to be domain specificity of response styles, as topic involvement is related to response styles (Diamantopoulos, Raeynolds, & Simintiras, 2006). More specifically, in the analysis of a large dataset of the Intenational Social Survey Program, van Dijk et al. (2009) tentatively concluded that response styles were more likely in domains of a high personal relevance compared to domains of a low personal relevance. The domain specific hypothesis is further examined in multiple datasets in the dissertation.

Effects of Response Styles in Cross-Cultural Studies

Response styles can influence both scale means and relationships with other variables (van de Vijver & Leung, 1997). There has been some controversy about the need to correct for response styles in survey research (D. H. Smith, 1967). On the one hand, the traditional interpretation that response styles present a distorted representation of participants' views makes correction imperative. An example of this approach can be found in Eysenck and Eysenck's (1975) work. They proposed to interpret scores on personality scales only if a participant's score on a social desirability scale was below a pre-determined critical threshold. On the other hand, various researchers have argued that corrections for response styles do not have a sizeable impact on conclusions based on self-report scores. For instance, Ones, Viswesvaran, and Reiss (1996) reported that the validity of an instrument to predict job performance is not strongly influenced by a correction for SDR.

In cross-cultural contexts, response styles are found to differ across cultural groups. For example, African Americans and Hispanics were found to exhibit higher ARS and ERS than European Americans (e.g., Marin, Gamba, & Marin, 1992). Baron-Epel and colleagues (2010) reported higher ARS and ERS in Arabs than Jews in Israel. Morren, Gelissen, and Vermunt (2012) found that first-generation immigrants tended to use more ARS and ERS

compared with second-generation immigrants. In a comparison of Chinese, Japanese, American, and Canadian students' responses on Likert-scale items, Chen, Lee, and Stevenson (1995) found that these Asian students had a higher score on MRS compared with students from the two North American countries. A similar conclusion was reached in Hamamura, Heine, and Paulhus (2008). Comparing response styles in six European countries, van Herk, Poortinga and Verhallen (2004) concluded that Southern Europe scored higher on ARS and ERS than Northwestern Europe.

Given the variations in response styles across cultures, there seems to be a need to adjust for response styles in order to have valid cross-cultural comparisons. Welkenhuysen-Gybels, Billiet, and Cambré (2003) demonstrated that omitting a factor accounting for ARS could lead to a biased assessment of the invariance of the target construct across groups. Tellis and Chandrasekaran (2010) found that response styles led to inaccurate conclusions about innovativeness based on survey data as compared to that based on the market penetration of new products in 15 countries.

However, the correction effects in cross-cultural studies have shown mixed results. Diamantopoulos, Reynolds, and Simintiras (2006) claimed that response styles had an inconsistent impact on cross-cultural differences. Correcting for response styles shifted the country ranking on substantive constructs in some domains, but in other domains there was no change. Chen et al. (1995) found that response styles did not alter cross-cultural comparisons of item means in a four-country comparative study. Dudley et al. (2005) reported that correcting for SDR did not affect the validity of a personality test in different racial groups. In an eight-country study, Hoffmann, Mai, and Cristescu (2013) reported very small changes in correlations and mean comparisons after adjusting for response styles. The inconsistent findings reviewed above may be attributed to the various operationalizations used to gauge response styles, and the specific samples and constructs of interest, which prevent us from generalizing these findings. Using a stable, integrated response style, this dissertation addresses the correction effects of response styles among different ethnic groups in the Netherlands and in large-scale international surveys.

Main Research Questions

I am interested in systematizing the measurement of response styles and unraveling the psychological meaning of response styles. In this dissertation, three main research questions are posed.

1. **Is there a General Response Style that can integrate specific response styles including ARS, ERS, MRS and SDR at both individual and cultural level?** In other words, can more consistency in measurement and findings of response styles be achieved, through the examination of the shared meaning of these specific response styles?

2. **What are the nomological network and cross-cultural variations of the General Response Style and each specific response style at both individual and cultural level?** Specifically, what are the correlates of response styles in domains such as socioeconomic development, values, personality, and political views, and whether there is domain dependency of response styles (i.e., personal relevance elicits more response styles)? Can measures of other formats than Likert-scale self-report measures that are robust to the effects of response styles shed light on the nature of response styles?

3. **What are the implications of response style (correction) effects in cross-cultural comparative studies?** That is, how does response style correction affect the effect size and group ranking in self-report measures in different ethnic groups in the Netherlands and in cross-national surveys? How does response style correction converge or diverge from other methods proposed in the literature to deal with scale usage differences including score standardization in the Schwartz Value Survey, overclaiming (i.e., respondents' tendency to self-enhance independent of their ability) and anchoring vignettes (i.e., a procedure to rescale respondents' self-report based on how they rate several hypothetical persons described in written vignettes on the same traits)?

Outline of the Dissertation

The empirical part of this dissertation includes nine separate chapters under three sections. It should be noted that each chapter was developed as an independent manuscript, thus they can be read separately.

Section 1 (Chapter 2 to Chapter 4) is entitled the Integration of Specific Response Styles among Ethnic Groups in the Netherlands. This section examines how the specific response styles from different samples and instruments can be integrated to a General Response Style, and how this General Response Style can be generalized to a self-presentation style in surveys (Research question 1). Chapter 2 focuses on the confirmation

of a General Response Style factor from both direct and indirect measures of specific response styles in five ethnic groups in the Netherlands. Chapter 3 extends Chapter 2 by replicating the General Response Style factor in a Dutch national representative sample and studying the domain dependency and correction effects of the General Response Style. Chapter 4 is a further investigation of the General Response Style. In this longitudinal study, the General Response Style is linked with the general factors derived from personality and value questionnaires, and all of them are indicators of self-presentation in surveys.

Section 2 (Chapter 5 to Chapter 7) is entitled Cross-Cultural Variations in Response Styles. This section studies the nomological network of the General Response Style and each specific response style with various validity measures (Research question 2). Chapter 5 deals with the dimensionality and cross-cultural variations in social desirability, using an adapted Marlowe-Crowne Social Desirability scale. Chapter 6 extends the individual-level General Response Style to country level. Large-scale international survey data are used to study response styles at both levels. Chapter 7 makes use of a forced-choice format personality measure across countries to validate the substantive meaning of the General Response Style and each specific response style.

Section 3 (Chapter 8 to Chapter 10) is entitled Implications of Response Styles for Cross-Cultural Score Differences. This section targets the effects of response styles and (other scale usage correction methods) on the validity of cross-cultural score differences (Research Question 3). Chapter 8 zooms in on the effects of acquiescence and social desirability in association with the Schwartz Value Survey and other international datasets, and in particular the effects of score standardization as a means to control for response styles. Chapter 9 investigates the effects of the General Response Style on estimates of cross-cultural differences in 18 countries with data from the 2013 Teaching and Learning International Survey; both domain dependency and the correction effects are addressed. Chapter 10 compares the effects of response styles and two other methods to correct for scale usage, namely overclaiming and anchoring vignettes, with data of the Programme for International Student Assessment (PISA).

In Chapter 11, a summary of the findings of the above mentioned empirical studies is discussed with an emphasis on the meaning and implication of response styles in cross-cultural contexts.

SECTION ONE

THE INTEGRATION OF SPECIFIC RESPONSE STYLES AMONG ETHNIC GROUPS IN THE NETHERLANDS

Chapter 2

A General Response Style Factor: Evidence from a Multi-Ethnic Study in the Netherlands

This chapter is based on
He, J., & van de Vijver, F. J. R. (2013). A General Response Style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, 55, 794-800. doi:10.1016/j.paid.2013.06.017

We are interested in response styles, defined as the systematic tendency to respond to questions on some basis other than the target construct (Paulhus, 1991). The most studied response styles include acquiescence (ARS: the tendency to agree regardless of item content), extremity (ERS: the tendency to overuse the end points of a scale), midpoint responding (MRS: the tendency to overuse the middle point of a scale), and socially desirable responding (SDR: the tendency to answer questions in a way that makes oneself look good). Although conceptually related, these four response styles are seldom studied simultaneously. Little is known about their similarities and differences. Furthermore, the psychological meaning of response styles is not clear. Two interpretations can be found in the literature. The first, conventional perspective holds that response styles are nuisance factors and should be avoided as much as possible (Hui & Triandis, 1989). The alternative view interprets response styles as communication styles, indicating that they have a substantive meaning and that they reflect culture-moderated communication filters (P. B. Smith, 2004). Such a filter could moderate or amplify responses, as usually found in East Asia and Latin America, respectively. Moreover, response styles are found to be closely related to personality traits. Unlike previous investigations that have focused on specific response styles, we aim to integrate the four response styles and study their commonalities and differences, the cross-ethnic variations, and the associations with personality traits in a multicultural context.

The Interrelatedness of Response Styles

The definitions and correlates of ARS, ERS, MRS, and SDR suggest that they are related. ERS, a tendency to be unequivocal with a self-promotion focus, can be viewed as the opposite of MRS, a tendency to be evasive with a prevention focus (van Vaerenbergh & Thomas, 2013). Smith and Fischer (2008) found that ARS was more salient among collectivists and ERS more among individualists. Negative associations between the two can be expected. SDR and ERS have in common that they represent desirable traits related to extroversion and conscientiousness (Austin et al., 2006; Musek, 2007). We expect that there is a single factor underlying these four response styles, with positive loadings of ERS and SDR, and negative loadings of ARS and MRS (Hypothesis 1). We do not expect this first factor to explain all covariation among the indicators, as previous research already suggested that each indicator has some uniqueness (P. B. Smith, 2011).

Cross-Ethnic Variations of Response Styles

It has been argued that immigrant groups, compared with the majority group, are under higher pressure not to deviate much from the general norm (Arends-Tóth & van de Vijver, 2009). African Americans and Hispanics were found to exhibit higher ARS and ERS than European Americans (e.g., Marin et al., 1992). Baron-Epel and colleagues (2010) reported higher ARS and ERS in Arabs than Jews in Israel. Morren, Gelissen, and Vermunt (2012) found that first-generation immigrants tended to use more ARS and ERS compared with second-generation immigrants. So, groups with a culture further away from the dominant group tend to show higher levels of ARS and ERS.

We argue that the differences in response style use among minority groups and the majority group may be a function of both perceived cultural distance and prevailing in-group values (Davis, Resnicow, & Couper, 2011). Comparing with the majority group, minority groups may tend to use more moderating communication strategies such as ARS and MRS in order to "fit in" the society. In addition, minority groups with a collectivistic background (typically from Non-Western cultures), who value loyalty to their cultural heritage and espouse allegiance to in-groups, may exhibit more moderating communication styles to demonstrate conformity to in-groups. In general, we expect more ARS and MRS use among minority groups with a larger cultural distance to the majority group and with a collectivistic orientation (Hypothesis 2).

Response Styles and Personality Traits

There is abundant evidence on the associations of response styles and the Big Five personality traits. For example, ARS and SDR have been found to be related to agreeableness (e.g., Graziano & Tobin, 2002); ERS and reversed MRS are positively related to extraversion (e.g., Austin et al., 2006); and the self-deceptive enhancement dimension of SDR is negatively related to neuroticism (Pauls & Stemmler, 2003). Beyond these specific effects, the "Big One" personality (i.e., general factor of personality) was found to be strongly related to SDR (e.g., Just, 2011), causing controversies in the substantive interpretation of the "Big One" personality. Irwing (2013) critically reviewed the multi-method multi-trait models and cross-validations of the general factor model, supporting that the "Big One" personality is unlikely to be a measurement artifact. We apply multiple measures to construct a General Response Style and expect a strong general effect of this

style on the "Big One" factor. In addition, we expect specific associations of specific response styles with specific personality traits.

The Present Study

There is a tradition to operationalize response styles as proportions of specific score patterns on usually heterogeneous sets of items, such as the endorsement of either extreme of a Likert scale as ERS (Paulhus, 1991). However, given the evidence that response styles are stable across time and throughout questionnaires , it should be possible to assess them directly; for example, one could ask self-reports about the importance of having a strong opinion as a measure of ERS. The present study addresses both conventional, indirect and direct self-reports of response styles. We aim to integrate the four response styles into one General Response Style factor and examine (1) their interrelatedness in direct and indirect modes; (2) cross-ethnic similarities and differences in response styles; and (3) their associations with personality traits.

We conducted the study in the Netherlands, where immigrants constitute 21% of the total population, from which 45% are of Western origins (e.g., European, North American), and 55% are of Non-Western origins (e.g., Turkish, Moroccan, Surinamese, Antillean). Around 50% are first-generation and 50% are second-generation immigrants (Statistics-Netherlands, 2011). These immigrant groups have different levels of similarity to the Dutch society. Generally, Non-Western immigrants are less similar than Western immigrants to Dutch nationals; first-generation immigrants are less similar compared with second-generation immigrants (Arends-Tóth & van de Vijver, 2009).

Method

Participants

In this paper use is made of immigrant panel data of the MESS (Measurement and Experimentation in the Social Sciences) project administered by CentERdata (Tilburg University, The Netherlands). The immigrant panel is a representative sample of Dutch immigrants and majority group members who participate in monthly Internet surveys. The panel is based on a true probability sample of households drawn from the population register. Households that could not otherwise participate are provided with a computer and Internet connection. In the present study, participants were 1664 panel members from five ethnic

groups: Dutch nationals, first- and second-generation immigrants of Western and Non-Western origins. The demographics of the sample are presented in Table 2.1.

Table 2. 1 ***Descriptive Characteristics of Participants in the Five Groups***

Ethnic Groups	Sample size	Mean age	% of males	Mean Education	Latent mean of GRS	Latent mean of mode factor
Dutch nationals	548	47.49	51%	3.79	0 (fixed)	0 (fixed)
2nd G Western	344	47.05	46%	3.59	.01 (.04)	.02 (.11)
1st G Western	253	51.78	42%	4.13	-.05 (.04)	.06 (.13)
2nd G Non-Western	173	31.32	40%	3.58	-.16 (.05)**	.33 (.15)*
1st G Non-Western	346	43.66	47%	3.62	-.10 (.04)**	.56 (.12)**

Note. Education was scored from 1 (*primary school*) to 6 (*university*). G = generation. GRS = General Response Style. Numbers between parentheses in last two columns refer to standard errors. *$p < .05$. **$p < .01$(two-tailed).

Measures

Indirect measures of ARS, ERS, and MRS. We extracted indirect measures of ARS, ERS, and MRS from data in the panel archive (http://www.lissdata.nl/). ARS was extracted from the scales of *Self-Esteem* and *Survey Attitude*, in total 16 items, both with half positively and half negatively worded items and with 7-point *disagree* to *agree* response options. The ARS score was operationalized as the proportion of the responses of 5 (*somewhat agree*) and 6 (*agree*). Responses of 7 (*strongly agree*) were excluded from the ARS computation due to the fact that such responses may also be triggered by ERS. We avoid the common problem that the correlation between ARS and ERS is overestimated, when the *strongly agree* responses are used to compute both ARS and ERS.

ERS was constructed from sets of 5-, 6-, and 7-point scales that used various response anchors (e.g., *not at all* to *very much so*, *extremely unimportant* to *extremely important*) other than *strongly disagree* to *strongly agree*. Item contents were heterogeneous, including affects, autobiographical memory, emotion, health, personality and values. The proportion of the two end point responses (e.g., 1 and 5 in the 5-point scale) was taken as the ERS score. We only use the odd-numbered items from the item pool (109 in total) for the indirect ERS.

MRS was constructed from the even-numbered items using 5- and 7-point response scales (85 in total) in the same data pool as ERS. The proportion of the midpoint responses (e.g., 4 in the 7-point scale) was taken as the MRS score. The three indirect measures were constructed in this way to avoid (1) confounding of response styles with the substantive

constructs of the data source, and (2) data dependency among the three response styles. The scores of ARS, ERS, and MRS ranged from 0 to 1, with a higher value indicating a higher level of the response style.

Self-report measures of ARS, ERS, MRS, and SDR. We developed and piloted self-report measures of ARS, ERS, and MRS. We used balanced scales with semantic differentials, which have been shown to enhance cross-cultural comparability and to induce fewer response styles (e.g., Friborg, Martinussen, & Rosenvinge, 2006). Specifically, the scales comprised questions with half positively and half negatively phrased items. All response formats had seven response anchors, which varied from item to item (e.g., *never* to *always*, *not important at all* to *extremely important*).

Principal Component Analysis revealed a one-factor structure for ARS (explained variance of 36.64%); a sample item from the 10-item scale read "Do you sometimes say "Yes" even though you do not actually agree?" The Cronbach's Alpha of the scale ranged from .80 to .82 across the five groups.

ERS items loaded on two factors and items on the first factor were taken (explained variance of 26.56%). A sample item from the 5-item ERS scale (α values ranged from .69 to .74) read "Do you like to be viewed as a person with strong opinions?"

All MRS items loaded on one factor (explained variance of 29.04%), and a sample item in the 10-item scale read "Do you prefer neutral opinions to strong opinions?" (α values ranged from .66 to .74).

We selected and simplified 17 items from the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) and the Balanced Inventory of Desirable Responding (Paulhus, 1991) to assess impression management (IM) (e.g., "I help others in trouble") and self-deceptive enhancement (SDE) (e.g., "I am confident about my judgment"). These two dimensions have been found important to understand the nature of SDR (Pauls & Stemmler, 2003). All the SDR items were rated on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Principal Component Analysis supported the two-factor solution, with explained variances of 18.27% and 13.20%. We deleted two items with cross-loadings. Values of Cronbach's Alpha ranged from .70 to .74 for IM (11 items), and from .58 to .66 for SDE (4 items).

To demonstrate the structural and scalar equivalence of the self-report response style scales, we carried out multigroup confirmatory factor analyses for each scale across the five

ethnic groups in AMOS (Byrne, 2001). We checked invariance of measurement weights (i.e., factor loadings on the latent variable were constrained to be equal across groups) and invariance of intercepts (i.e., items were constrained to have the same intercepts across groups). The model fit was evaluated by Chi-square tests, the Tucker Lewis Index (acceptable above .90), Comparative Fit Index (acceptable above .90), and Root Mean Square Error of Approximation (acceptable below .06) (G. W. Cheung & Rensvold, 2002). The same criteria apply to the subsequent analysis on the General Response Style factor. For each self-report response style scale, invariance of measurement weights was supported, and in most cases the fit decreased when invariance of intercepts was taken into consideration (Table 2.2). In all, the fit of these intercepts invariance models were acceptable and we concluded that scalar equivalence of the scales was fairly well supported.

Table 2. 2 ***Measurement Invariance of the Self-Report Response Style Scales***

Scale	Invariance	χ^2/df	TLI	CFI
Acquiescence	MW	1.85**	.94	.95
	Intercepts	2.06**	.93	.92
Extremity	MW	3.88**	.91	.94
	Intercepts	2.95**	.94	.93
Midpoint Responding	MW	2.21**	.89	.91
	Intercepts	2.48**	.86	.87
Impression Management	MW	1.90**	.90	.91
	Intercepts	1.96**	.89	.89
Self-Deceptive Enhancement	MW	1.37**	.99	.99
	Intercepts	1.84**	.97	.96

Note. TLI = Tucker–Lewis index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; MW = measurement weights. **$p < .01$ (two tailed).

Personality measures. Data of the Big Five personality traits collected among the same respondents in a previous wave were used. The scores of *Agreeableness*, *Conscientiousness*, *Extroversion*, *Openness*, and *Neuroticism* were taken from the 50-item International Personality Item Pool (Goldberg et al., 2006) with response options ranging from 1 (*very inaccurate*) to 5 (*very accurate*). The reliability of the five traits was high (values of α ranged from .70 to .89). We obtained the "Big One" personality score though factor analyzing the five personality traits (Musek, 2007). Principal Components Analysis revealed a one factor solution (with explained variance of 40.11%), which had positive loadings of Agreeableness (.68), Conscientiousness (.61), Extroversion (.70), Openness (.70), and a negative loading of Neuroticism (-.45).

Results

We describe the results in three parts: the fit of the General Response Style model, the tests of the mean differences of response styles, and the associations of response styles with personality.

The General Response Style Factor Model

We fitted a General Response Style factor model in a multigroup confirmatory factor analysis across the five ethnic groups. We tested the model with all the direct and indirect response style measures loading on the General Response Style factor; all the self-report measures having identical loadings and all the indirect measures having another set of identical loadings on an additional mode factor (uncorrelated with the General Response Style factor), which was to account for the effect of data collection method. The structural covariance model was the most restrictive model with an acceptable fit (Table 2.3).

Table 2. 3 ***Results of the Multigroup Confirmatory Factor Analysis: The General Response Style***

Model	χ^2/df	*df*	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf
Unconstrained	2.97**	83	.91	.95	.03		
Measurement weights	2.62**	111	.93	.94	.03	44.48*	28
Measurement intercepts	2.88**	143	.91	.91	.03	120.36**	32
Structural covariances	*2.82***	*147*	*.92*	*.91*	*.03*	*3.20*	*4*
Measurement residuals	2.85**	191	.91	.88	.03	130.12**	44

Note. Most restrictive model with acceptable fit is printed in italics. TLI = Tucker–Lewis index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.
*$p < .05$. **$p < .01$ (two-tailed).

The standardized solution of the model is presented in Figure 2.1. We found support for a General Response Style factor encompassing both direct and indirect measures, with positive loadings of ERS and the two subscale of SDR, and negative loadings of ARS and MRS. Hypothesis 1 was confirmed. The General Response Style factor explained 27.83% of the variance, suggesting that there was considerable overlap in response styles, although clearly not all variation was captured.

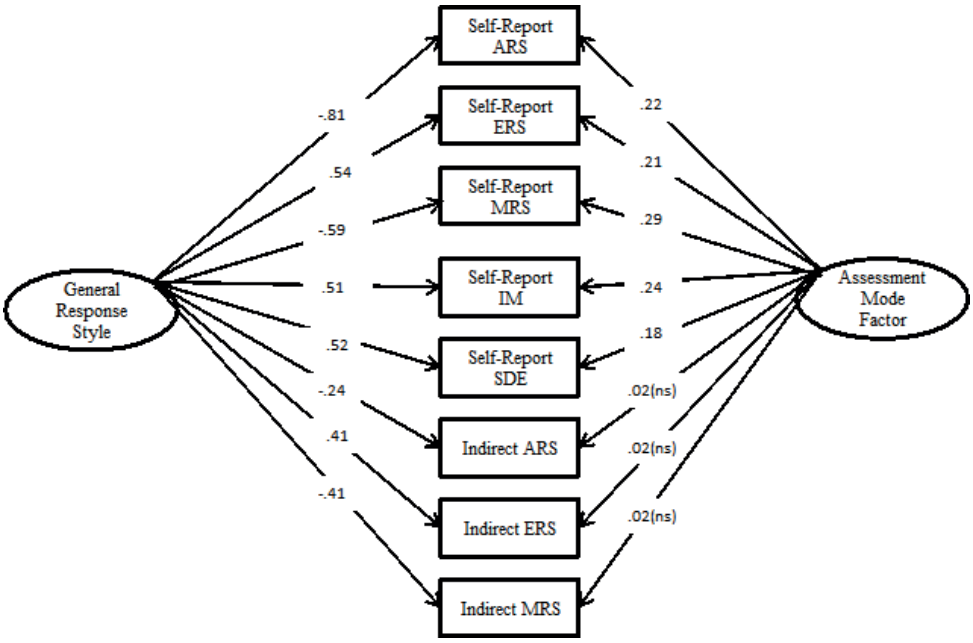


Figure 2. 1 Standardized Solutions of the General Response Style Model from Acquiescence (ARS), Extremity (ERS), Midpoint Responding (MRS), Impression Management (IM), and Self-Deceptive Enhancement (SDE). All coefficients were significant at $p < .01$(two-tailed), except these specified as nonsignificant (*ns*).
Note. The values of the constrained loadings on the assessment mode factor were slightly different due to the standardization procedure. There were three correlated error terms not shown in the figure, which were between self-report ERS and IM, indirect ARS and MRS, and indirect ERS and MRS.

Cross-Ethnic Variations of Response Styles

We first compared the latent means of the General Response Style and the mode factor, fixing the means of both factors for the Dutch national group to zero in the measurement intercepts model. The model fitted reasonably well, $\chi^2(135, N = 1664) = 360.20$, $p < .01$, TLI = .92, CFI = .93, and RMSEA = .03. We found that both generations of Non-Western immigrants had lower means of the General Response Style factor and higher means of the mode factor compared with Dutch nationals, whereas both generations of Western immigrants did not significantly differ from Dutch nationals (Column 5 and 6 in Table 2.1). Because only the self-report measures had significant loadings on the assessment mode factor, the mean differences on this factor might indicate that the distortion in self-reports was higher in the Non-Western immigrant groups compared with the other groups.

We examined group differences in specific response styles in a MANCOVA, with all the response style scores as dependent variables, group membership as the independent

variable, and age, gender, and education as covariates. The overall multivariate test was significant, Wilks' Lambda = .94, $F(32, 6016) = 3.16$, $p < .01$, $\eta^2 = .02$; however, the mean differences among the groups were very small (Table 2.4). We found that immigrants, especially first-generation Non-Western immigrants, scored higher on self-report ARS, both measures of MRS, and self-report SDE than Dutch nationals. Age, gender, and education had significant effects with small to medium effect sizes ($\eta^2 = .09$, .08, and .04, respectively). The General Response Style factor (scored in the direction of ERS and SDR versus ARS and MRS) correlated positively (all *p*s < .05) with age [$r(1662) = .19$], gender (coded as 0 for female and 1 for male) [$r(1662) = .08$], and education [$r(1644) = .06$].

Table 2. 4 ***Comparison of Response Style Scale Means among the Five Groups in MANCOVA (with age, gender, and education as covariates)***

	Dutch National	2nd G Western	1st G Western	2nd G Non-Western	1st G Non-Western	F	η^2
Self-report ARS	2.84	2.80	2.95	2.90	2.97	2.46*	.01
Self-report ERS	4.87	4.95	4.90	4.97	4.85	0.88	.00
Self-report MRS	3.63_{a}	3.62_{a}	$3.66_{a,b}$	$3.73_{a,b}$	3.81_{b}	4.75**	.01
Self-report IM	5.57_{a}	$5.54_{a,b}$	5.39_{b}	$5.54_{a,b}$	5.60_{a}	2.95*	.01
Self-report SDE	$4.77_{a,b}$	$4.77_{a,b}$	4.64_{a}	$4.70_{a,b}$	4.93_{b}	3.04*	.01
Indirect ARS	.34	.34	.34	.36	.34	0.55	.00
Indirect ERS	.24	.24	.24	.22	.25	0.95	.00
Indirect MRS	$.21_{a}$	$.21_{a}$	$.23_{a,b}$	$.23_{a,b}$	$.24_{b}$	6.67**	.02

Note. G = Generation. ARS = Acquiescence, ERS = Extremity, MRS = Midpoint Responding, IM = Impression Management, SDE = Self-Deceptive Enhancement. Means were estimated marginal means when age, gender, and education were corrected for; means with different subscripts are significantly different in Bonferroni post hoc comparison; Degrees of freedom = 4, 1658. *$p < .05$. **$p < .01$ (two-tailed).

Associations with Personality Traits

The General Response Style, and most of the specific response styles, showed moderate to strong correlations with the Big Five personality traits (Table 2.5). The relationships between response styles and personality traits were explored in a conceptual model where response styles served as communication filters to the expressions of personality traits. We modeled the effect of the General Response Style factor to the "Big One" factor in a multigroup path model with latent factors. We expected to find a general effect from the response style factor to the "Big one" factor, and given the large amount of unexplained variance in the factor analysis reported before, we also expected some specific

Table 2. 5 ***Correlations of Response Styles with Personality Traits***

Personality	GRS	Self-report ARS	Self-report ERS	Self-report MRS	Self-report IM	Self-report SDE	Indirect ARS	Indirect ERS	Indirect MRS
Agreeableness	.27	-.14	.12	-.09	.39	.10	-.10	.36	-.34
Conscientiousness	.33	-.21	.17	-.11	.32	.28	-.15	.34	-.30
Extroversion	.37	-.27	.35	-.28	.12	.25	-.12	.23	-.19
Openness	.35	-.27	.30	-.22	.14	.17	-.12	.28	-.34
Neuroticism	-.35	.31	-.07	.17	-.25	-.35	.19	-.24	.24
"Big One"	.52	-.37	.34	-.27	.37	.35	-.20	.46	-.45

Note. GRS = the General Response Style, ARS = Acquiescence, ERS = Extremity, MRS = Midpoint Responding, IM = Impression Management, SDE = Self-Deceptive Enhancement. All correlations are significant ($p < .01$; two-tailed).

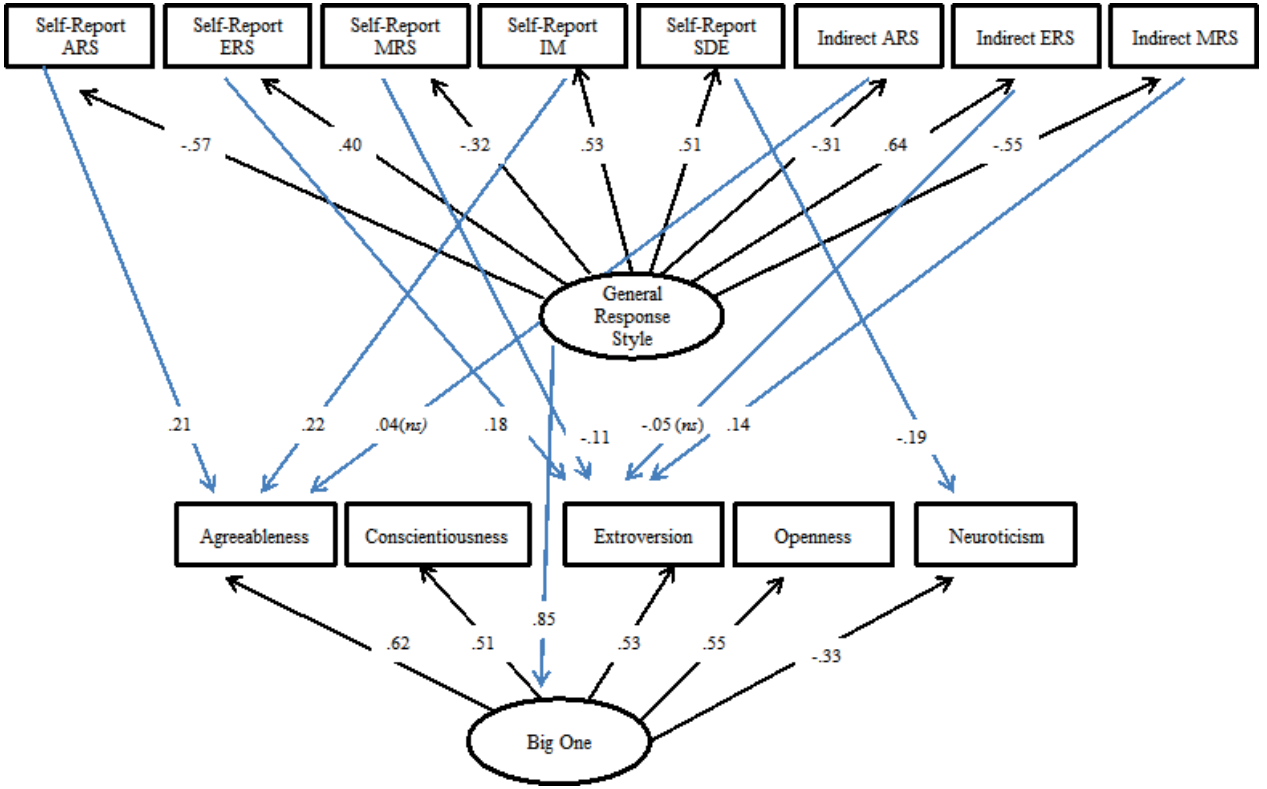


Figure 2. 2 Standardized Solutions of the Associations among the personality traits and the General Response Style Factor derived from Acquiescence (ARS), Extremity (ERS), Midpoint responding (MRS), Impression Management (IM), and Self-Deceptive Enhancement (SDE) among the Five Groups.

Note. All coefficients were significant at $p < .01$ (two-tailed), except these specified as nonsignificant (*ns*).We allowed correlated error terms among the self-report response styles and among the indirect measure of response styles to model mode differences.

relationships between personality and specific response styles. We found a marginally acceptable fit for a structural residual model ($\chi^2(328, N = 1301) = 798.16$, $p < .01$, TLI = .87, CFI= .89, and RMSEA = .03), in which the General Response Style factor strongly predicted the "Big One" personality (standardized $\beta = .85$) and the specific response styles predicted specific personality traits in a multigroup analysis (Figure 2.2). Self-report ARS and IM predicted agreeableness, ERS and MRS predicted extroversion, and SDE predicted neuroticism. These correlations are largely in line with the expectations from the literature.

Discussion

We examined four response styles among five ethnic groups in the Netherlands. We found support for a General Response Style factor which differed across the groups, and which had a strong association with personality traits. Our study contributes to the literature in two ways. First, the confirmation of a general factor of response styles delineates a new avenue to conceptualize and analyze response styles. Second, we used multiple sources of data, collection methods, and samples to assess response styles, which provided more robust estimates.

Our study has implications for the discussion about the meaning of response styles as nuisance or substance. We showed that the different response styles have both shared and unique aspects. If response styles are viewed as communication styles, our study shows that individuals have a tendency to moderate or amplify their responses (the general response factor), that is related to expressing themselves on personality questionnaires, as evidenced in the strong correlation with the "Big One" factor (Irwing, 2013). In addition, each response style has a unique component that is meaningfully related to specific personality traits, such as the link between extroversions and ERS. If response styles are viewed as nuisances, our study shows that the nuisance has both a general component (the response style factor) and specific components (e.g., ERS is particularly challenging when assessing extroversion). Response styles are then a complex set of confounding variables threatening personality assessment. However, referring to these styles as a nuisance does not enhance the understanding of their nature. Rather, we argue that our study shows how the two views can be integrated by referring to response styles as communication filters that influence how individuals express themselves when filling out personality questionnaires (e.g., P. B. Smith, 2004). Therefore, we argue our response style factor can be tentatively interpreted as

referring to impression management strategies that act as "communication filters" in expressing oneself. The filter can be assessed both directly and indirectly and has various indicators such as ERS and MRS. Its close association with the "Big One" personality factor suggests that both may refer to the combination of positive personality states (Just, 2011; Musek, 2007), and both may be a high-level indicator of the preferable way of how people want to present themselves to others.

We found small and fairly inconsequential cross-cultural differences in response styles across ethnic groups in the Netherlands. The variation in the mode factor suggests possible self-report distortion: Non-Western immigrants exhibit more distortion than Western immigrants and Dutch nationals. The integration of response style data from multiple measures will provide a more stable estimate and may make findings with regard to response styles more stable.

Limitations and Future Directions

Our study has a few limitations. First, we extracted a single ARS, ERS, and MRS index from items mainly dealing with private life domain questions. Future efforts should construct indexes with various formats from various life domains to ensure that the effects we found are not domain or format specific. Second, we used Likert-type self-reported measures of personality traits. It would be interesting to have other measures that are supposed to be less susceptible to response styles (e.g., forced-choice questions), or other techniques to effectively control for response styles (e.g., anchoring vignettes). Third, we confirmed the model of response styles (i.e., communication filters) impact traits in both general and specific ways in a cross-sectional design, yet longitudinal and experimental studies should be carried out to substantiate the relationship. In conclusion, a new line of investigation opens up with the integration of specific response styles, and it is worth researching to enlarge our insight in the nature of response styles.

Chapter 3

Integration and Domain Specificity of Response Styles: Towards a Better Understanding of a General Response Style

This chapter is based on
He, J., & van de Vijver, F. J. R. (2014). Integration and domain specificity of response styles: Towards a better understanding of a General Response Style.

Submitted for publication

Response styles are defined as the systematic tendency to use certain categories of an answering scale on some basis other than the target construct (Cronbach, 1950). It is imperative to understand the underlying mechanism of response styles, thus appropriate measures can be taken to address their effects in self-reports. The most frequently studied response styles include acquiescent response style (ARS), extreme response style (ERS), midpoint response style (MRS), and socially desirable response style (SDR) (Paulhus, 1991). Although usually treated as distinct, these four response styles are related. It has been found that a General Response Style (GRS) can be extracted with ERS and SDR as positive indicators, and ARS and MRS as negative indicators; GRS is strongly associated with desirable personality traits and can be interpreted as a communication filter (i.e., response amplification to moderation) (He & van de Vijver, 2013). However, associations of GRS with presumably relevant constructs (e.g., values and self-regulation) have not been adequately researched, nor has its impact on these psychological measures been explored. In the present study we first address the replicability of GRS with a national representative sample. We then address novel topics, including the nomological network of response styles and the effects of their correction.

Nomological Network of Response Styles

Response styles are associated with various psychological measures, notably personality. ERS and SDR were positively related to extroversion and conscientiousness (Eid & Rauber, 2000; Meiser & Machunsky, 2008). ARS and the impression management dimension of SDR were related to agreeableness (Messick & Jackson, 1961; Pauls & Stemmler, 2003). Regarding values, the impression management dimension of SDR was positively associated with collectivism and the self-deceptive enhancement dimension with individualism (Lalwani et al., 2009). Regarding self-regulation, ERS was positively related to promotion regulation focus and MRS to prevention regulation (Cabooter, 2010). In the cognition domain, ERS was positively associated with intolerance of ambiguity (Naemi et al., 2009), suggesting that respondents' need to evaluate enhances the use of ERS. Krosnick (1991) argued that respondents avoiding cognitive efforts in survey responding tend to utilize ARS as a satisficing strategy. Positive life outcomes, such as subjective well-being and self-esteem, were positively related to SDR and ERS (e.g., Bachman & O'Malley, 1984). These correlations can be extended to GRS, based on the loadings of these specific response styles on GRS.

Domain Specificity of Response Styles

Topic involvement boosts response styles (Diamantopoulos et al., 2006). ERS was found to be higher in surveys dealing with personal domains than public domains (van Dijk et al., 2009). It seems that in self-evaluation respondents' need for controllability is higher and the motivation to self-enhance or self-protect is more salient. We expect these findings to extend to GRS. The correlations of GRS with other self-report constructs may differ according to the level of self-involvement, as domains of higher self-involvement (e.g., personality) may share more variance with GRS, compared with domains of lower self-involvement (e.g., political views).

Correcting for Response Styles

Response styles can influence both scale means and relationships with other variables (van de Vijver & Leung, 1997). The conventional interpretation of response styles as nuisance factors suggests that score corrections will increase scale validity; yet, the effectiveness of response style correction is not a foregone conclusion. It was argued that ERS and MRS could both hide and exaggerate cross-national differences, yet their correction effects are inconsistent (Diamantopoulos et al., 2006). Dudley et al. (2005) reported that correcting for SDR did not affect the validity of a personality test in different racial groups. Inconsistencies in findings may be due to differences in response styles across stimulus formats, life domains, and operationalizations of response styles. Part of the inconsistencies could presumably be accommodated by integrating specific response styles to a GRS. We explore effects of GRS correction on the patterning of correlations of variables from different domains. If GRS indeed serves as a communication filter that is equally reflected in all self-report measures, GRS correction on any self-report measure would result in consistently lower intercorrelations among these variables. If GRS influences self-report measures of various domains differentially, changes vary across domains. Finally, we explore how the correction of GRS could impact the size of cross-group differences on target measures.

Method

Participants and Procedure

In this paper use is made of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). The LISS panel is a representative sample of Dutch individuals who participate in monthly Internet surveys. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Households that could not otherwise participate are provided with a computer and Internet connection. In the present study, respondents of three cultural groups were sampled: Dutch mainstreamers (n = 4,812; mean age = 50.63 years; male 46%), Western immigrants from other European countries and North America (n = 395, Mean age = 52.48 years; male 46%), and Non-Western immigrants from Turkey, Morocco, Suriname and other Eastern countries (n = 250; Mean age = 39.96 years; male 44%).

Measures

Indirect measures of ARS, ERS, and MRS

We constructed ARS, ERS, and MRS indexes using non-overlapping sets of items to avoid data dependency (the items used and the syntax to construct these response style indexes are available from the first author upon request).

An ARS index was constructed from eight bi-directional scales (i.e., scales with positively and negatively worded items). These scales covered various life domains (e.g., optimism, attitude to public drug policy). All scales had 5- or 7-point response options with anchors from *strongly disagree* to *strongly agree*. We averaged the proportions of the agreeing categories (e.g., scores of 4 in the 5-point scales, and scores of 5 and 6 in the 7-point scales) from the positively worded items and the negatively worded items for the ARS index. The endorsement of *strongly agree* was not taken as a part of ARS, to avoid confounding with ERS. The value of Cronbach's alpha from the 77 items (recoded as 1 for ARS endorsement and 0 for Non-ARS endorsement, respectively) was .71.

An ERS index was extracted from sets of 4-, 5-, and 7-point scales in the Likert-scale item pool with various item contents. We randomly selected 15 items from each response format (i.e., 4-, 5-, and 7- point response options), recoded the responses as ERS endorsement (e.g., scores of 1 and 4 of the 4-point items as 1) and Non-ERS endorsement (e.g., scores of 2 and 3 of the 4-point items as 0), and averaged the ERS endorsement from the 45 items as an indicator of global ERS (α = .88).

Similar to the ERS index, an MRS index was constructed from sets of 5- and 7-point Likert-scale items. Fifteen items were randomly chosen from each response format and the proportion of the mid-point endorsement (e.g., scores of 4 of the 7-point items) was taken as the global MRS index ($\alpha = .64$).

Self-report measures of ARS, ERS, MRS, and SDR

Self-reported measures of ARS, ERS, and MRS have been developed and validated in He and van de Vijver (2013), and the same measures were used in the present study. For each style, a 10-item scale comprised questions with half positively and half negatively phrased items, and each item had a 7-point response option with anchors varying from item to item.

Principal component analysis revealed a one-factor structure for ARS (explained variance of 38%); a sample item of the scale read "Do you sometimes say "Yes" even though you do not actually agree?" ($\alpha = .81$).

The 10 ERS items loaded on two factors; the first factor was about expressiveness and preference of ERS, whereas the second factor was more about the functional aspects of utilizing ERS. Items on the first factor were taken (explained variance of 30%). A sample item from the 5-item ERS scale read "Do you like to be viewed as a person with strong opinions?" ($\alpha = .72$).

All the MRS items loaded on one factor (explained variance of 28%), and a sample item in the MRS scale read "Do you prefer neutral opinions to strong opinions?" ($\alpha = .69$).

Seventeen items from the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) and the Balanced Inventory of Desirable Responding (Paulhus, 1991) were administered to tap into impression management (IM) (e.g., "I help others in trouble") and Self-Deceptive Enhancement (SDE) (e.g., "I am confident about my judgment"). All the SDR items were rated on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Principal component analysis supported the two-factor solution, with explained variances of 24% and 10% ($\alpha = .73$ for IM; $\alpha = .63$ for SDE).

We ensured the equivalence of all the self-report response style measures across the three cultural groups with multigroup confirmatory factor analysis. For each self-report response style scale, a model testing invariance of measurement weights and intercepts showed an acceptable fit, which indicated that the scale means can be compared directly.

Nomological network measures

We used data of personality, values, self-regulation, cognition, positive life outcomes, and political views administered to the same participants in the panel.

Personality. Fifty items of the International Personality Item Pool (Goldberg et al., 2006) were administered to assess the Big Five personality factors. Response options ranged from 1 (*very inaccurate*) to 5 (*very accurate*). The "Big One" personality factor was also calculated as the factor score of the five traits in a principal component analysis, in which all the five traits loaded positively on the factor (Musek, 2007). The values of Cronbach's alpha for the five traits were .81, .79, .87, .77, and .88, respectively.

Values. *Horizontal Individualism* (sample item: "I'd rather depend on my own strength than being dependent on others"), *Vertical Individualism* ("Winning is everything"), *Horizontal Collectivism* ("If an acquaintance gets a prize, I would feel proud"), and *Vertical Collectivism* ("It is important to me that I respect the decisions made by my groups") were assessed with an 8-item scale from Triandis and Gelfand (1998). The horizontal-vertical dimension distinguishes the preferences of equality versus hierarchy. The response options ranged from 1 (*totally not applicable*) to 7 (*totally applicable*). The values of Cronbach's alpha were .76, .76, .57, and .61, respectively.

Self-regulation. Two emotion regulation strategies were assessed with the 10-item scale from Gross and John (2003) with response options ranging from 1 (*strongly disagree*) to 7 (*strongly agree*): *Reappraisal* (i.e., tendency to change an emotional response by reinterpreting the meaning of the emotional stimulus) (α = .81) and *Suppression* (i.e., tendency to inhibit ongoing emotion-expressive behavior) (α = .78).

Two regulation focus measures, *Prevention Focus* (i.e., focus on the avoidance of failure) (α = .85) and *Promotion Focus* (i.e., focus on the pursuit of success) (α = .90), were taken from the scale from Lockwood, Jordan, and Kunda (2002); response options ranged from 1 (*totally not applicable*) to 7 (*totally applicable*).

Cognition. *Need to Evaluate*, defined as the extent to which people spontaneously evaluate objects or experiences as either good or bad, was measured by the 16-item scale by Jarvis and Petty (1996). The response options ranged from 1 *(extremely uncharacteristic of me*) to 5 (*extremely characteristic of me*). The 16 items loaded on two factors and only the 10 items that loaded on the first factor were used to calculate the scale score (α = .85).

Need for Cognition, representing the extent to which people engage in and enjoy a wide variety of effortful cognitive activities, was measured by the 18-item scale by

Cacioppo and Petty (1982). The response options ranged from 1 (*strongly disagree*) to 7 (*strongly agree*) (α = .89).

Positive outcomes. *Satisfaction with Life* was measured by the 5-item scale from Diener, Emmons, Larsen, and Griffin (1985). The response options ranged from 1 (*totally disagree*) to 7 (*totally agree*) (α = .90). *Self-esteem* was measured by the 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1965), with response options ranging from 1 (*totally disagree*) to 7 (*totally agree*) (α = .90).

Political views. *Opinion on Democracy* (sample item: "voters should have direct input in law-making") was measured with 9 items (α = .78), *Trust in Politicians* ("elected politicians are the most suitable people to make decisions") was measured with 4 items (α = .65), and *Opinion on Immigration* ("the unity of the Netherlands is weakened by the immigrant population") was measured with 10 items (α = .88). The response options for the three scales ranged from 1 (*totally disagree*) to 5 (*totally agree*).

Results

We report the results in three parts: the fit of the GRS model across cultural groups, the associations of response styles with the nomological network measures, and the impact of GRS correction on the correlation pattern and mean differences.

The General Response Style Factor Model

We fitted the General Response Style model in a multigroup confirmatory factor analysis in AMOS (Byrne, 2001), in which indirect measures of ARS, ERS, and MRS, and self-report measures of ARS, ERS, MRS, IM, and SDE all loaded on the General Response Style factor; all the indirect and direct measures also loaded on an assessment mode factor with identical loadings per mode (to account for the difference in data collection modes). The measurement residuals model was the most parsimonious model that fitted well (Table 3.1).

The standardized solutions are presented in Figure 3.1. We found support for a General Response Style factor with positive loadings of ERS and SDR, and negative loadings of ARS and MRS. Only the self-report measures loaded significantly on the mode factor, which indicates that the mode factor served to capture the distortion in self-report measures.

Table 3. 1 ***Results of the Multigroup Confirmatory Factor Analyses of the General Response Style***

	χ^2	*df*	χ^2/df	TLI	CFI	RMSEA	$\Delta\chi^2$	Δdf
Unconstrained	651.41	*49*	13.29**	.88	.93	.05		
Measurement weights	671.65	63	10.66**	.91	.93	.04	20.24	14
Measurement intercepts	729.67	79	9.24**	.92	.92	.04	58.02**	16
Structural covariances	733.95	81	9.06**	.92	.92	.04	4.28	2
Measurement residuals	*827.87*	*103*	*8.04***	*.93*	*.92*	*.04*	*93.91***	*22*

Note. Most restrictive model with acceptable fit is printed in italics. TLI = Tucker–Lewis index. CFI = Comparative Fit Index. RMSEA = Root Mean Square Error of Approximation. ***p* < .01.

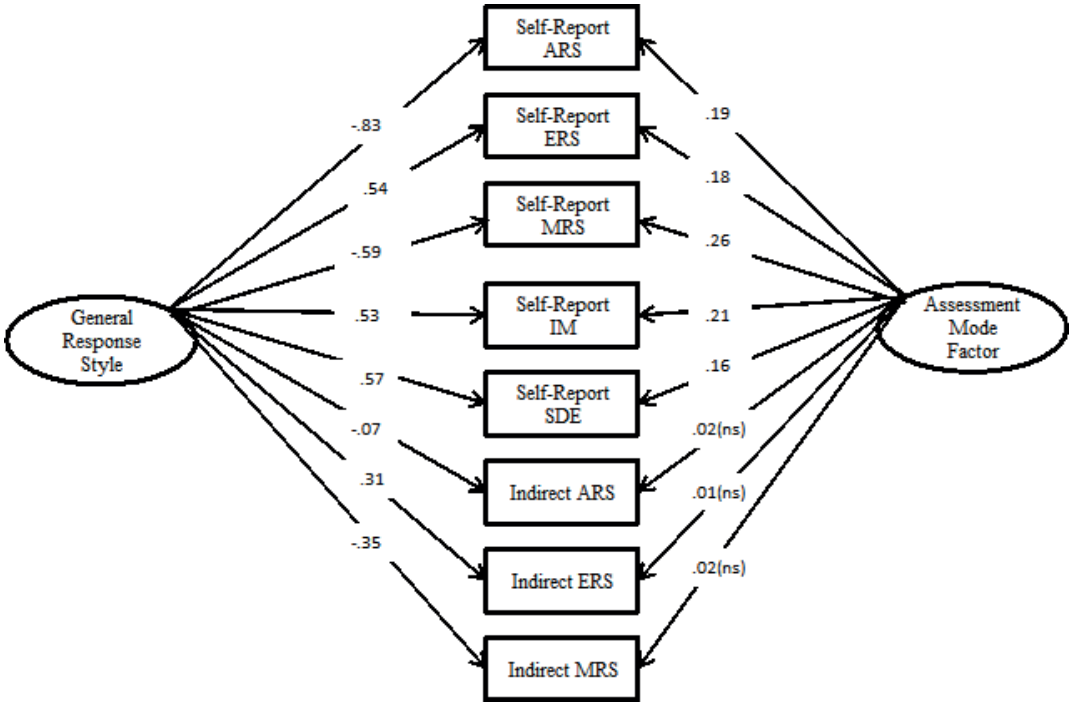


Figure 3. 1 Standardized Solutions of the General Response Style Model from acquiescent response style (ARS), extreme response style (ERS), midpoint response style (MRS), Impression Management (IM), and Self-Deceptive Enhancement (SDE) among the three ethnic groups. All coefficients were significant at $p < .01$, except that the loadings of indirect measures on the mode factor were nonsignificant (*ns*).
Note. The values of the constrained loadings on the assessment mode factor were slightly different due to the standardization procedure. There were three correlated error terms not shown in the figure, which were between self-report ERS and IM, indirect ARS and indirect ERS, and indirect ERS and indirect MRS.

We compared the latent means of GRS and the mode factor across ethnic groups, using the measurement intercept model that fitted reasonably well ($\chi^2(75, N = 5{,}457) = 700.98$, $p < .01$, TLI = .92, CFI = .93, RMSEA = .04). Compared with Dutch mainstreamers (as the reference group), Non-Western immigrants scored lower on GRS ($M = -.08$, $SE = .04$, $p < .05$), whereas Western immigrants scored higher ($M = .06$, $SE = .03$, $p < .05$). Non-Western immigrants had a higher mean on the mode factor ($M = .52$, $SE = .13$, $p < .01$), whereas the Western immigrant group did not differ from the Dutch mainstream group ($M = .13$, $SE = .10$, $p > .05$). These differences were in line with He and van de Vijver (2013).

Domain Specificity and Meaning of Response Styles

The correlations of response styles with all the other psychological variables are presented in Table 3.2. GRS showed moderate to strong positive associations with the desirable personality traits, horizontal individualism, cognition, and self-esteem, and a negative association with prevention self-regulation focus. Political views had very weak associations with GRS. Overall, the correlations with personality and cognition were stronger than with other domains, which implies that measures that were more self-evaluative triggered more GRS. The domain specificity of GRS was supported.

The specific response styles had similar, yet weaker correlation patterns than GRS. ERS and MRS seemed to be the essential indicators of GRS, as they had a very similar patterning to GRS. There was more differentiated patterning in ARS and the two SDR dimensions. ARS was particularly related to prevention focus and trust in politicians. Both SDE and IM were related to collectivism which stresses interpersonal harmony; SDE was also positively related to vertical individualism.

Impact of GRS Correction

The scores of each of the other psychological variables were predicted by GRS, and the standardized residuals of each variable were saved as GRS-corrected scores. We compared the intercorrelations among variables using raw scores and GRS-corrected scores. In the personality domain (five traits and the "Big One" personality factor), the absolute positive correlations among them dropped on average by .10 (from .36 in the raw scores to .26 in the corrected scores), indicating that the GRS shared quite some variance with personality traits. The correlation between the two cognition measures dropped by .09 (from .32 to .23), and between the two positive outcome measure by .06 (from .47 to .41).

Table 3. 2 ***Correlation of Response Styles with Other Variables***

	GRS	ARS direct	ARS indirect	ERS direct	ERS indirect	MRS direct	MRS indirect	SDE	IM
Extraversion	.37**	-.29**	-.02	.29**	.20**	-.24**	-.25**	.23**	.11**
Agreeableness	.28**	-.10**	-.04**	.09**	.23**	-.07**	-.32**	.09**	.40**
Conscientiousness	.37**	-.19**	-.06**	.15**	.31**	-.09**	-.32**	.28**	.34**
Emotional Stability	.42**	-.35**	.00	.12**	.28**	-.17**	-.23**	.39**	.29**
Openness	.38**	-.29**	.04*	.32**	.15**	-.28**	-.35**	.19**	.11**
"Big One" Personality	.58**	-.39**	-.03	.32**	.37**	-.28**	-.48**	.37**	.39**
Horizontal Individualism	.32**	-.24**	.00	.25**	.17**	-.14**	-.23**	.21**	.17**
Vertical Individualism	-.01	.01	.12**	.17**	-.12**	-.01	-.01	.09**	-.18**
Horizontal Collectivism	.17**	-.07**	.06**	.12**	.08**	-.02	-.20**	.13**	.21**
Vertical Collectivism	.13**	-.04**	.05**	.09**	.06**	.02	-.15**	.10**	.18**
Reappraisal	.12**	-.07**	.02	.08**	.10**	.02	-.13**	.07**	.11**
Suppression	-.11**	.05**	.04*	-.06**	-.01	.14**	.13**	.01	-.10**
Prevention Focus	-.33**	.30**	.10**	-.04*	-.31**	.20**	.11**	-.24**	-.24**
Promotion Focus	-.02	.07**	.12**	.19**	-.18**	.01	-.09**	.04*	-.14**
Need to Evaluate	.37**	-.30**	.06**	.50**	.17**	-.26**	-.19**	.22**	.01
Need for Cognition	.32**	-.29**	.04*	.29**	.06**	-.29**	-.22**	.18**	.07**
Life Satisfaction	.24**	-.12**	.08**	.05**	.15**	-.06**	-.21**	.30**	.20**
Self-Esteem	.55**	-.41**	-.08**	.23**	.33**	-.23**	-.35**	.47**	.37**
Opinion on Democracy	.06**	-.04*	-.03	.08**	.09**	.05**	.02	.09**	.03
Trust in Politicians	-.04*	.07**	.11**	-.03	-.06**	.07**	-.07**	.04*	.01
Opinion on Immigration	-.05**	.06**	-.05**	.03	.03	.09**	.05**	.02	-.09**

Note. GRS = General Response Style. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. IM = Impression Management. SDE = Self-Deceptive Enhancement.
*$p < .05$. **$p < .01$.

Table 3. 3 ***Results of MANOVA***

Group	Type of Scores	Wilks' Lambda	*df*1	*df*2	Partial η^2
Age cohort	Raw Scores	.80**	40	3106	.11
	GRS-corrected	.81**	40	3106	.10
Gender	Raw Scores	.78**	20	1554	.22
	GRS-corrected	.78**	20	1554	.22
Education	Raw Scores	.64**	100	7556	.08
	GRS-corrected	.65**	100	7556	.08
Cultural group	Raw Scores	.90**	40	3106	.05
	GRS-corrected	.89**	40	3106	.06

Note. Age cohorts include three groups: age below 35, age 35 to 65, and age above 65. Gender was scored as 0 for female and 1 for male. Education was scored from 1 (*primary school*) to 6 (*university*). Cultural Group refers to Dutch Mainstreamers, Western Immigrants, and Non-Western Immigrants.
**$p < .01$.

The average change (in absolute values) in the self-regulation domain was .02, that of value domain .01, and that of political views was .00. To summarize, the effects of GRS on the relationship between variables were small, yet meaningfully patterned. Its influence was in line with the personal involvement level of domains: the larger the self-evaluative component, the larger the correction. In whatever domain, the change of correlation was weak at best.

A series of MANOVAs were performed with all the other psychological variables (first in raw scores and then with GRS correction) as dependent variables and demographic variables as independent variables; the effect sizes of these analyses were compared (Table 3.3). Specifically, we compared the partial eta squared values of the differences in 21 dependent variables from both raw and GRS corrected scores across three age groups, between males and females, across the six levels of education, and across the three cultural groups. The effect sizes of group differences were largely identical in the two sets of scores. We examined the change of effect sizes per domain and did not find any significant difference in the sizes. However, it should be noted that even with raw scores, there were very limited cross-cultural mean differences in these variables, which was possibly due to the insufficient heterogeneity in our samples (e.g., the immigrant participants in the panel are well adjusted to the Dutch society).

Discussion

Our study is, to our knowledge, the first to address the combined meaning of ARS, ERS, MRS, and SDR with multiple measures in a national representative sample. We found a GRS with positive loadings of ERS and SDR and negative loadings of ARS and MRS. This general factor has stronger associations with personality traits, cognition, and positive life outcomes than with other domains. The impact of GRS on the structure of self-report data was in accordance with a domain dependence model: GRS correction affects domains with high personal involvement more than domains with low personal involvement; yet, even in personal domains, score corrections did not have a strong impact on correlations or gender, age and ethnic differences.

The confirmation of GRS illustrates that specific response styles can be integrated, and the combined meaning represents the general tendency to amplify or to moderate responses. GRS, as well as the specific response styles, is not associated with various self-

report measures to the same degree, but vary across domains. It seems that response styles have more influence on responses in domains with more self-involvement (van Dijk et al., 2009). We also observed a near-zero correlation of GRS with vertical individualism, possibly for the same reason. Whereas all the other items within the value domain start with the first-person pronoun "I", items on vertical individualism are formulated in a less personal, more distant manner (e.g., "winning is everything"; "competition is the law of nature").

Integrating different response styles in GRS produces more reliable estimates of response styles, as it reconciles various problems resulting from different operationalizations of response styles, or the domain differences when extracting indirect measures of specific response styles from other self-report measures. Furthermore, our findings suggest reluctance in response style correction. Firstly, we observed differential sizes of correlations between response styles and various psychological measures. These styles share variance notably with personality measures. However, our results indicate that the global impact of score corrections on correlations is modest and there is no evidence for a validity increase after correction (Ones et al., 1996). Secondly, the correction of GRS does not change the structure of self-report measures. All in all, GRS may represent valid individual and cross-cultural differences, and it can be a strong candidate to define an important aspect of culture: a preferential way of self-disclosure in communication (P. B. Smith, 2011).

Limitations and Future Directions

A few caveats should be mentioned. We did not have external measures that are less susceptible to response styles (e.g., behavior measures) to validate GRS-corrected scores. Besides, due to the lack of mean differences across the cultural groups in the current study, we could not draw conclusion on the effects of GRS on the magnitude of scale means across groups; replicating the study with heterogeneous cross-cultural samples which presumably have a larger mean difference may shed light on this issue. In conclusion, different response styles can be studied simultaneously with the extraction of GRS. However, our study suggests GRS should be more the topic of future research in stable communication preferences rather than a source of score correction to increase the validity of self-reports.

Chapter 4

Self-Presentation Styles in Self-Reports: Linking the General Factors of Response Styles, Personality Traits, and Values in a Longitudinal Study

This chapter is based on
He, J., & van de Vijver, F. J. R. (in press). Self-presentation styles in self-reports: Linking the general factors of response styles, personality traits, and values in a longitudinal study. *Personality and Individual Differences* (Young Researcher Special Issue). doi: 10.1016/j.paid.2014.09.009

Most social interactions, including attitude expression and change, can be analyzed through the lens of self-presentation (Baumeister, 1982). Self-presentation is the use of behavior as a means of communicating information about (or an image of) oneself to others. Depending on personal dispositions and specific contexts, people exhibit preferred self-presentation styles, such as being assertive or defensive, acquisitive or protective, and active or passive. Response styles, defined as respondents' systematic tendencies to respond to questionnaires on some basis other than the target constructs (Paulhus, 1991), can be considered an essential indicator of self-presentation in survey responses (e.g., P. B. Smith, 2004). We are interested in response styles that affect self-report Likert scales in different domains; more specifically, we set out to examine whether individuals show a similar self-presentation style across measures of personality (i.e., personal style) and values (i.e., normative function), and to what degree this style changes over time. The novelty of our study is that we use a variety of psychological measures to investigate self-presentation across psychological domains in a longitudinal study.

Response Styles and Their Integration

The most studied response styles include acquiescent response style (ARS), extreme response style (ERS), midpoint response style (MRS), and socially desirable responding (SDR). Traditionally, these response styles are viewed as sources of common method bias that should be controlled for (Podsakoff & Organ, 1986). Recent evidence, however, suggests that response styles may have a substantive meaning, as they are found to share trait variance with personality and values. For example, ERS was positively associated with extroversion, conscientiousness, and individualistic values, SDR with extroversion and conscientiousness, ARS with agreeableness and compliance, and MRS with modesty (e.g., Austin et al., 2006; Chen et al., 1995). To integrate these findings, He and van de Vijver (2013) confirmed that these four response styles could be integrated into a General Response Style (GRS), with ERS and SDR as positive indicators and ARS and MRS as negative indicators. This integrated GRS is interpreted as a preferred communication style that represents the tendency of response amplification versus moderation.

Response Styles, Personality, and Values

J. A. Johnson (1981) suggested that response styles and the expressions of personality and values share some commonality, possibly all related to one's self-presentation styles. In line with this suggestion, the general factor of personality derived from the Big Five personality traits, interpreted as a basic personality disposition integrating the most general

non-cognitive dimensions of personality (Musek, 2007), was found to be positively correlated with GRS (He & van de Vijver, 2013). Bye et al. (2011) reported that personal values, associated with ARS, were related to intended impression management. If a general factor of values were to be extracted (e.g., Podsakoff & Organ, 1986), we expect this factor to be indicative of individuals' self-presentation styles.

Stability and Change in Self-Presentation Styles

Little has been done on the stability and changes of self-presentation styles. Yet, the stability of response styles over time has been demonstrated. Weijters, Geuens, and Schillewaert (2010) used different item sets at two time points to measure response styles over a one-year period. They modeled time-invariant and time-specific response style factors, and found that response styles showed considerable stability. We expect longitudinal stability of self-presentation styles. Furthermore, in this study items common to all data collection points are used as indicators of response styles, and expressions of personality and values, thus both the stability and the changes of their integration over time could be tested.

The Present Study

The literature suggests that response styles and the expressions of personality and values may all be part of self-presentation styles. The stability and changes of self-presentation styles over time, in turn, may affect the associations among self-report measures. We used a general factor of response styles derived from socially desirable, extreme, and midpoint responding, a general factor of personality based on the Big Five personality traits, and a general factor of values based on five value dimensions across three time points (T1, T2, and T3), and we extracted a time-specific self-presentation factor from these three general factors at each time point to model its stability over time.

The effects of changes in the self-presentation factor over time could be demonstrated through comparing correlations of this factor with external psychological measures and intercorrelations among these external psychological measures with and without this factor corrected for at each time point. Well-established measures, such as self-esteem, life satisfaction, and positive and negative affect, were used as external measures. Researchers found that self-esteem, life satisfaction, and positive affect are positively related to each other, whereas negative affect shows a negative association with self-esteem and life satisfaction (Preisendörfer & Wolter, 2014; Robins, Fraley, Roberts, & Trzesniewski, 2001). If individuals indeed prefer a similar style across domains and time, we expect that the

correction of the self-presentation factor would result in similar changes (or lack of changes) in correlations with external variables and intercorrelations among external variables over time.

Table 4. 1 ***Demographics and Scale Properties at Each Time Point***

	T1	T2	T3
Demographics			
***N* of Participants**	6766	6980	6734
Mean Age (*SD*)	45.89 (15.95)	47.51(17.30)	48.94 (17.54)
Level of Education (Percentage)			
Primary school	11	12	12
Intermediate secondary education	26	26	25
Higher secondary education	10	11	11
Intermediate vocational education	24	22	22
Higher vocational education	22	22	22
University	7	7	8
Percentage of Males	46	46	46
Reliability of Scales (Cronbach's α)			
Extreme Response Style	.81	.82	.82
Midpoint Response Style	.57	.63	.66
Socially Desirable Responding	.52	.52	.52
Agreeableness	.80	.80	.80
Conscientiousness	.77	.79	.78
Extroversion	.86	.86	.87
Emotion Stability	.79	.79	.80
Openness	.77	.76	.77
Prosocial Concern	.90	.90	.90
Self-Directed Competence	.78	.79	.78
Restrictive Conformity	.81	.82	.81
Universal Maturity	.90	.90	.90
Stimulation/Comfort	.79	.80	.80
Self-Esteem	.89	.89	.90
Life Satisfaction	.88	.88	.89
Positive Affects	.87	.87	.87

Method

Sample and Procedure

In this paper use is made of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). The LISS panel is a representative sample of Dutch individuals who participate in monthly Internet surveys. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Households that could not otherwise

participate are provided with a computer and internet connection. A longitudinal survey is fielded in the panel every year, covering many domains.

We used five waves of data on value and character traits collected from 2008 to 2012, in which measures of the same constructs including affect, cognition, mood, personality, survey attitude, self-esteem, social desirability, trust, and values, in total 183 items, were administered. Each year, over 8,000 selected household members were invited to participate, and the numbers of respondents ranged from 5,321 to 6,806 (response rates ranging from 69.9% to 79.6%). In 2010 and 2012, the complete questionnaire was only administered to non-respondents of the previous wave; therefore we used data in 2008 as T1, combined data in 2009 and 2010 as T2, and those in 2011 and 2012 as T3. The demographics of respondents in the consolidated three time points are presented in the upper panel of Table 4.1.

Respondents who participated at all three time points (n = 3,879) were older (M = 51.43 years, SD = 15.71) than those who only took part in one or two time points (M = 44.38, SD = 18.36), $t(7,569) = 17.91$, $p < .01$, Cohen's $d = .41$. The education level differed slightly, the majority (49.5%) of those who participated at all three time points had an intermediate or higher vocational education level compared with 41.7% of those who participated once or twice. The difference of gender distribution between the two groups was nonsignificant, $\chi^2(1, N = 7,571) = .07$, $p = .80$.

Measures

Response styles. Response style indexes were derived from a wide variety of items. Three response styles were targeted: SDR, ERS, and MRS, because they were shown to be the defining indicators of GRS (He & van de Vijver, 2013). ARS was not included due to its ambiguous meaning due to the various operationalizations used in the past. When operationalized as the endorsement of agree proportions including the positive end (e.g., *strongly agree*), ARS is confounded with ERS; excluding the endorsement of the positive end results in an unclear meaning in ARS.

SDR. SDR was measured with a shortened version of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) with response options of 1 (*True*) and 0 (*False*). The reliabilities of this scale and of all the other measures at each time point are presented in the lower panel of Table 1. As can be seen there, values were adequate for all scales at each time point.

ERS. De Beuckelaer, Weijters and Rutten (2010) recommended that at least 15 items of heterogeneous content should be used to derive valid and reliable response style indexes.

Indexes of ERS and MRS were extracted from nonoverlapping items with 7-point Likert anchors from various measures in the questionnaire other than the Big Five personality and the Rokeach value items, in order to avoid data dependency between indexes and with the substantive measures of personality and values. Specifically, the same 15 items at the three time points were chosen to construct an ERS index. The average inter-item correlations ranged from .07 to .08, indicating sufficient heterogeneity in item content. The original responses were recoded as ERS endorsement (i.e., scores of 1 and 7 of the 7-point items as 1) and Non-ERS endorsement (i.e., scores of 2, 3, 4, 5, and 6 as 0), and the ERS endorsement from the 15 items was then averaged as an indicator of ERS.

MRS. A similar procedure was employed for an MRS index. Another 15 items were chosen from the item pool (average inter-item correlations ranging from .05 to .06), and recoded as MRS endorsement (i.e., scores of 4 of the 7-point responses as 1) and Non-MRS endorsement (i.e., scores other than 4 as 0). The average endorsement was taken as the index of MRS. The items used and the SPSS syntax to construct these indexes are available from the first author upon request.

A General Response Style factor (GRS), explaining 47% of all the variance, was extracted in a principal component analysis of the three response styles across all time points. As expected, ERS (.82) and SDR (.41) loaded positively and MRS (-.76) loaded negatively on the factor.

Personality. Fifty items of the International Personality Item Pool (Goldberg et al., 2006) were administered to assess the Big Five personality. Responses ranged from 1 (*very inaccurate*) to 5 (*very accurate*).

The general personality factor was extracted in a principal component analysis of the dimension, rather than item, scores of the five traits (*Agreeableness*, *Conscientiousness*, *Extroversion*, *Emotional Stability*, and *Openness*) across all time points (e.g., Musek, 2007). All five traits loaded positively on the factor (loadings ranging from .48 to .68), and it explained 38% of the variance.

Values. A rating-format version of the Rokeach Value Survey, including 18 instrumental values (i.e., preferred modes of behaviors) and 18 terminal values (i.e., desirable end-state of existence), was administered (Rokeach, 1973). The responses ranged from 1 (*extremely unimportant*) to 7 (*extremely important*). Following Feather (1991), we carried out a principal component analysis on the instrumental and terminal value items, respectively. Instead of using within-subject standardized scores, as proposed by Schwartz (1992) to account for response styles in value surveys, we used raw scores to analyze the expression of

values. The 18 instrumental values loaded on three factors: *Prosocial Concern* (e.g., sincere and truthful, responsible, and helpful), *Self-Directed Competence* (e.g., capable, independent, and creative), and *Restrictive Conformity* (e.g., hard-working, obedient, self-controlled), explaining 44%, 9%, and 6% of the variance, respectively. Two factors were extracted from the 18 terminal values: *Universal Maturity* (e.g., a world at peace, freedom, and inner harmony) and *Stimulation/Comfort* (e.g., social recognition, an exciting life, and a conformable life), explaining 41% and 12% of the variance, respectively.

Dimension, rather than item, scores of the five value dimensions across all time points were input into a principal component analysis to extract the general value factor. With all the value dimensions loading positively on the factor (loadings ranging from .77 to .85), it explained 66% of the variance.

External psychological measures. *Self-esteem* was measured by the 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1965), with responses ranging from 1 (*totally disagree*) to 7 (*totally agree*).

Life Satisfaction was measured by the 5-item Satisfaction with Life scale from Diener, Emmons, Larsen, and Griffin (1985). The responses ranged from 1 (*totally disagree*) to 7 (*totally agree*).

Positive and Negative Affects were measured with the 20-item PANAS Scale (Watson, Clark, & Tellegen, 1988) with responses ranging from 1 (*not at all*) to 7 (*extremely*).

Results

We describe the results in three parts. Firstly, we report the longitudinal measurement invariance of the three general factors (GRS, the "Big One" personality factor, and the general value factor), a prerequisite for valid longitudinal comparisons. Secondly, we report the modeling of the self-presentation factor from these general factors, and its stability and changes over time. Lastly, we describe effects of correcting for the self-presentation factor on the relationships among external psychological variables at each time point.

Longitudinal Measurement Invariance of the General Factors

To demonstrate the measurement equivalence of GRS, the "Big One" personality factor, and the general value factor over time, confirmatory factor analyses were carried out in AMOS (Byrne, 2001). For GRS, we specified three time-specific GRS factors (so, one factor per time point), where ERS, MRS, and SDR at each time point were indicators; these time-

specific GRS factors were correlated with each other, and uniqueness of each indicator was correlated across time points (e.g., ERS at T1, T2 and T3 were correlated) (configural invariance). We then added constraints to test invariance of measurement weights (i.e., factor loadings on GRS were equal across time points) and invariance of intercepts (i.e., indicators had the same intercepts across time points). The same model specifications were applied to the "Big One" personality factor and the general value factor. The model fit was evaluated by chi-square tests, the Tucker Lewis Index (acceptable above .90), the Comparative Fit Index (acceptable above .90), and the Root Mean Square Error of Approximation (acceptable below .06). The acceptance of a more restricted model was based on the change of CFI (less than .01) (G. W. Cheung & Rensvold, 2002).

Due to 31% missing values of all the items across all time points, we resorted to full information maximum likelihood estimation in confirmatory factor analyses (Schafer & Graham, 2002). For each general factor, invariance of measurement weights and intercepts was well supported by the fit indexes (Table 4.2), although the chi-square values were rather high, which might be caused by the large sample size. We concluded that the longitudinal measurement invariance for the three general factors was confirmed, thus these factors could be compared across time points.

Table 4. 2 ***Measurement Invariance of the Scales: Measurement Weights and Intercept Invariance***

Scale	Invariance	$\chi^2(df)$	TLI	CFI	RMSEA
General Response Style[1]	Configural	222.89 (21)**	.97	.99	.03
	MW	248.51 (27)**	.98	.99	.03
	Intercepts	315.22 (33)**	.98	.98	.03
"Big One" Personality	Configural	1587.43 (72)**	.96	.97	.05
	MW	1598.66 (80)**	.96	.97	.04
	Intercepts	1914.35 (90)**	.96	.97	.05
General Value Factor[2]	Configural	37.8.37 (69)**	.91	.95	.07
	MW	3734.89 (77)**	.92	.95	.07
	Intercepts	4076.98 (87)**	.92	.94	.07

Note. [1]The variance of the error term for extreme response style was fixed to 0 to avoid negative variance.
[2]The error terms of Prosocial Concern and Stimulation/Comfort were negatively correlated.
TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = Root Mean Square Error of Approximation; MW = Measurement Weights. $^{**}p < .01$ (two tailed).

Stability and Change in Self-Presentation Style

We tested the model depicted in Figure 1 with full information maximum likelihood estimation. We used the factor scores of GRS, the "Big One" personality factor, and the

general value factor as observed variables. Each of the general factors of the preceding time points predicted that of the following time points, and a latent self-presentation factor was defined by the three general factors at each time point; their loadings were constrained to be equal across time (e.g., same loading of GRS on the self-presentation factor at the three time points). The model fitted well, $\chi^2(19, N = 9{,}935) = 96.66$, $p < .01$, TLI = .99, CFI = 1.00, and RMSEA = .02. The standardized regression weights and the explained variance for each endogenous variable are presented in Figure 4.1. All three observed general factors loaded positively on the self-presentation factor at each time point, indicating that participants had a similar style across personality and value domains, and that this self-presentation factor here represented amplifying versus moderating responses. The factor loadings of the general factors at T1 (.43 to .63) were larger than these of T2 and T3 (.25 to .40), which suggests a reduced impact of the self-presentation style across measurement occasions.

Due to model identification issues, the changes of the latent mean of the self-presentation factor over time could not be estimated in the model shown in Figure 1. We tested it instead in a longitudinal measurement invariance model, in which the three observed general factors loaded on the self-presentation factor at each time point. The measurement weights invariance model ($\chi^2(28, N = 9{,}935) = 1{,}572$, $p < .01$, TLI = .90, CFI = .94, RMSEA = .08) and intercept invariance model ($\chi^2(34, N = 9{,}935) = 1{,}948$, $p < .01$, TLI = .90, CFI = .92, RMSEA = .08) were largely supported. The latent means of the self-presentation factor were compared in the measurement intercept model with the mean of T1 fixed to zero. The model showed an acceptable fit, $\chi^2(32, N = 9{,}935) = 1{,}884$, $p < .01$, TLI = .90, CFI = .93, RMSEA = .08. All three indicators loaded positively on the latent factor (ranging from .22 to 1.00). Compared with T1, the self-presentation factor at T2 had a lower mean ($M = -.05$, $SE = .01$, $p < .01$), and that at T3 even lower ($M = -.10$, $SE = .01$, $p < .01$), pointing to the decrease of self-presentation over time.

Correction for the Self-Presentation Factor

Scores on self-esteem, life satisfaction, and positive and negative affects were correlated with the self-presentation factor (i.e., factor scores estimated in the model depicted in Figure 4.1) and with each other before and after the self-presentation factor was partialed out at each time point (Table 4.3). Only respondents for whom data at all three points were available were included in the correlation analyses ($n = 3{,}879$). The correlations of these external variables with the self-presentation factor were consistent across time points. Before correction, self-esteem and life satisfaction showed positive correlations with positive affect, and they had negative correlations with negative affect at all three time points. After partialing

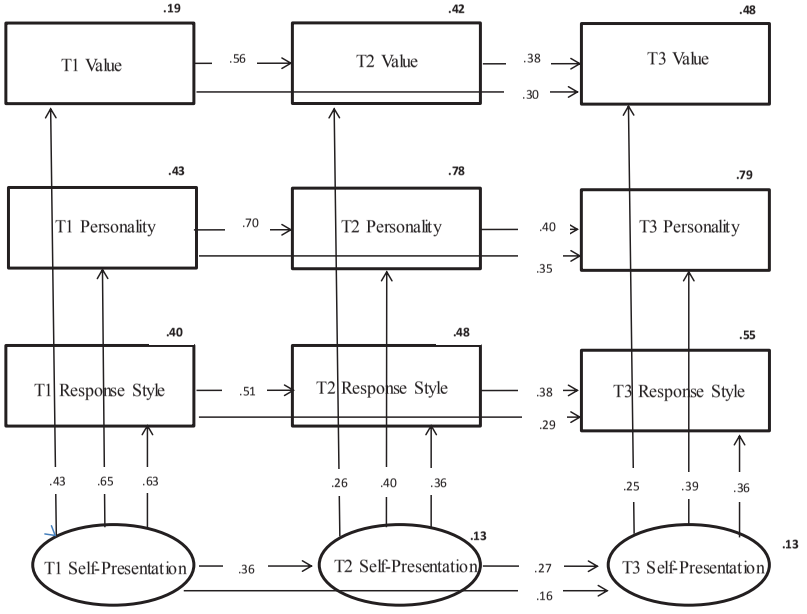


Figure 4. 1 Longitudinal Model of the Self-Presentation Style
Note. Standardized regression weights and factor loadings (all significant at $p < .01$) are presented next to the arrows. Numbers in bold represent proportions of variance explained.

out the self-presentation factor, the patterning of all correlations remained the same but the values became slightly weaker. The average absolute correlation dropped from .30 to .21 at T1, from .31 to .25 at T2, and from .32 to .26 at T3. All in all, the correction for the self-presentation factor across time does not seem to strongly affect the associations among these external variables.

Table 4. 3 ***Correlations with the Self-Presentation Factor and Among External Measures: Zero-Order Correlations (Below Diagonal) and With the Self-Presentation Factor Partialed Out (Above Diagonal) at Each Time Point***

	Self-Presentation Factor	1	2	3	4
T1					
1. Self-Esteem	.55	-	.32	.10	-.29
2. Life Satisfaction	.37	.45	-	.13	-.18
3. Positive Affect	.42	.31	.27	-	.26
4. Negative Affect	-.48	-.48	-.32	*.01*	-
T2					
1. Self-Esteem	.49	-	.41	.17	-.35
2. Life Satisfaction	.29	.49	-	.15	-.24
3. Positive Affect	.35	.30	.23	-	.20
4. Negative Affect	-.42	-.48	-.33	*.03*	-
T3					
1. Self-Esteem	.48	-	.40	.18	-.37
2. Life Satisfaction	.29	.48	-	.17	-.26
3. Positive Affect	.27	.29	.24	-	.19
4. Negative Affect	-.43	-.50	-.36	.04	-

Note. All correlations are significant at $p < .01$, except the ones italicized.

Discussion

We studied response styles, personality, and values from the perspective of an integrated self-presentation factor in a longitudinal study. We found that the general factors of response styles, personality, and values can be taken as indicators of a (global) self-presentation style. This self-presentation style and each general factor showed moderate to strong stability over time, suggesting that all of them may be part of a stable personal disposition. The loadings of the three general factors on the self-presentation factor and the latent mean of the self-presentation factor decreased over time. However, the impact of the self-presentation factor on changes in relationships among various psychological variables over time was very moderate.

We confirmed a stable self-presentation factor that is embedded in self-report data of various domains. In line with the interpretation of the GRS, which integrated specific response styles including ERS, SDR and MRS (He & van de Vijver, 2013), the self-presentation style represents amplifying (i.e., assertive and active) versus moderating responses (i.e., defensive and passive) to communicate one's image to others. As its indicators, the general factors of response styles, personality, and value load positively, suggesting that people tend to use the same style across different domains in survey responses, and more importantly that the self-presentation style shares variance with personality and values. Therefore, self-presentation style should not be merely interpreted as a domain specific nuisance factor in surveys.

We also found that the latent mean of the self-presentation factor decreased somewhat over time, suggesting that self-presentation becomes less salient over time. It is unlikely that the changes are caused by a similar decrease in values and personality over time; instead it suggests that the changes are due to scale usage. With repeated administrations of the same self-evaluative questions, the self-presentation style seems to become less salient, probably because the apprehension of performing an unfamiliar task (at T1) declines as the task becomes more familiar (at T2 and T3) (Ackerman, 1987), or responses over time are less influenced by cultural norms to present oneself in a culturally accepted way. The self-presentation factor affects all the psychological variables that we considered in the study; yet, the decrement of this factor does not strongly affect its correlations with external variables, nor the intercorrelations among these variables, which further suggests that self-presentation is not an independent nuisance factor but is integrated in the assessment of target constructs. Like the GRS, the first factors emerged from personality and value measures seem to reflect respondents' scale usage. There has been much debate on the nature of the general personality factor, and our study confirms findings by He and van de Vijver (2013) that it is part of a self-presentation style that is not merely an artifact.

We do not concur with the suggestion to control for the effects of self-presentation in survey responses. The loadings of the response styles indicate strongly the presence of individual differences in a style of responding, which can be influenced by personal preference as well as inculcated cultural values, notably conformity (He, van de Vijver, Domínguez, & Mui, 2014). Such a response style is part and parcel of one's psychological makeup and cannot be easily teased out. Therefore, statistical removal of response styles (part of self-presentation style) is unlikely to increase the validity of scores. Moreover, the stability of the self-presentation style construct over time (though combined with a mean change) and

the stability of its impact on various psychosocial measures over time also speak against the effectiveness of applying a temporal separation of measurements (i.e., introduce a time lag between the measurements of target variables) as a means to control for response styles (Weijters et al., 2010).

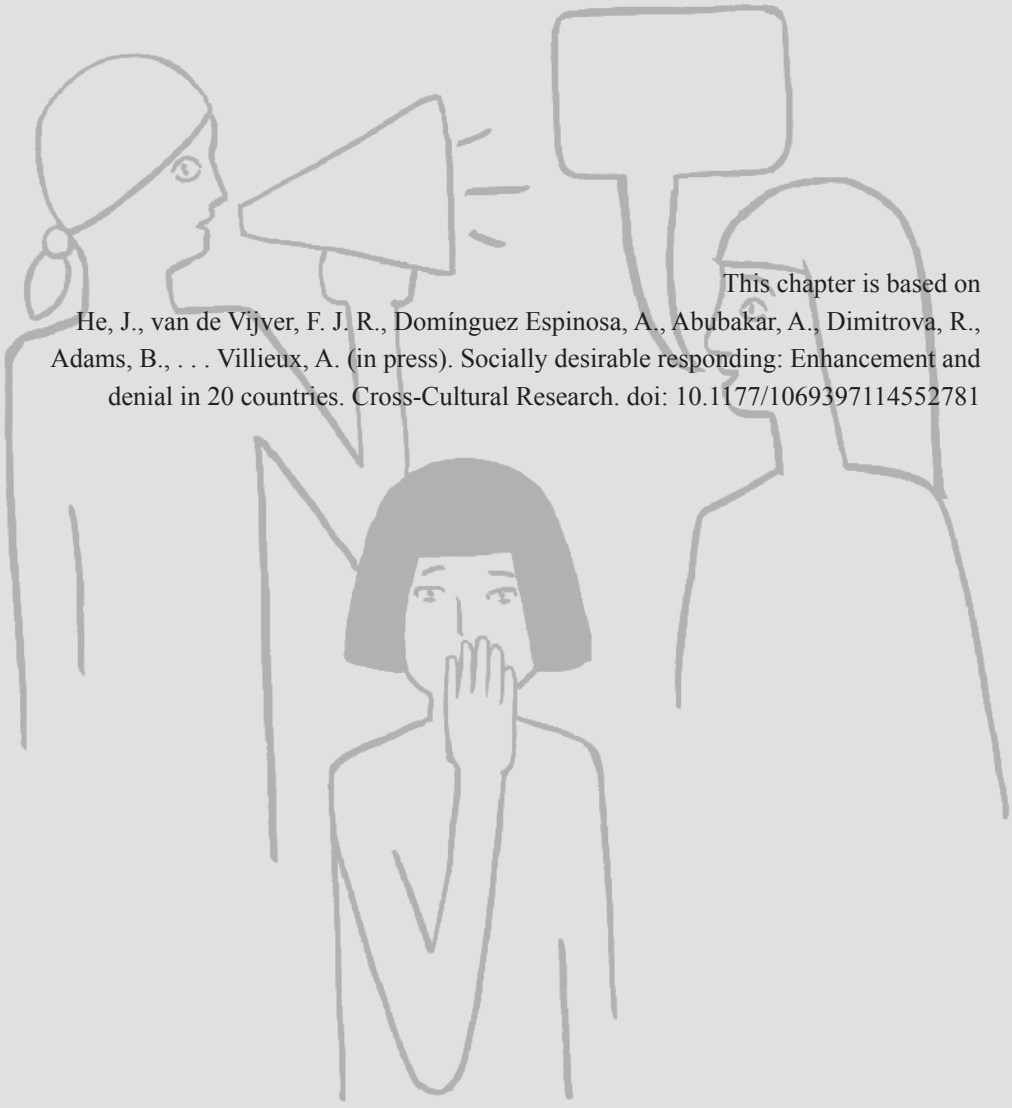
In conclusion, our study demonstrated the pervasiveness of self-presentation in response styles and self-report personality and values, and its stable, yet weak effect on the associations of self-reports. In spite of a few limitations in our study such as the inevitable information loss due to attrition in longitudinal data and the omission of acquiescence, our findings inform researchers about a general self-presentation style affecting all kinds of self-reports possibly due to social norms, and the caution needed for any score correction in self-reports.

SECTION TWO

CROSS-CULTURAL VARIATIONS IN RESPONSE STYLES

Chapter 5

Socially Desirable Responding: Enhancement and Denial in 20 Countries



This chapter is based on
He, J., van de Vijver, F. J. R., Domínguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B., . . . Villieux, A. (in press). Socially desirable responding: Enhancement and denial in 20 countries. Cross-Cultural Research. doi: 10.1177/1069397114552781

Socially desirable responding (SDR) refers to the tendency of respondents to reply in a manner that will be viewed favorably by others (Paulhus, 1991). SDR can challenge the validity of psychological measures. Respondents high on SDR tend to respond according to how they think people in their immediate environment would like them to react, whereas the interpretation of psychological measures is based on responses that are not contaminated by SDR. In this line of thinking, SDR is a nuisance factor that should be minimized, through a careful research design or statistical corrections (e.g., Nederhof, 1985). In another interpretation, SDR is more about substance than style (Schwartz, Verkasalo, Antonovsky, & Sagiv, 1997); it is then considered part and parcel of the psychological makeup of individuals that reflects culturally preferred ways of communication associated with various other cultural characteristics (van Hemert, van de Vijver, Poortinga, & Georgas, 2002). Here, SDR reflects genuine individual and cultural differences, representing effective and truthful self-presentation. The accurate measurement of this construct is a prerequisite for resolving the nuisance versus substance interpretation of SDR (Leite & Beretvas, 2005). However, the steadily accumulating empirical evidence about probably the most widely used SDR instrument, the Marlowe-Crowne scale (Crowne & Marlowe, 1960), has not produced clear conclusions as to its dimensionality and cross-cultural equivalence (e.g., Li & Reb, 2009; Verardi et al., 2010). Not surprisingly, the debate of the nature of SDR is continuing and a clear psychological meaning of SDR has not been adequately established. In our view, cross-cultural evidence could help to examine the stability of its dimensionality and provide important information about the nuisance versus substance discussion. With a shortened and adapted version of the Marlowe-Crowne scale, the present study examines the factor structure of SDR across 20 countries, and the associations of SDR with country-level characteristics.

Dimensionality of the Marlowe-Crowne scale

Studies of SDR do not reveal the same number of factors and different instruments yield very different factor structures (Paulhus, 2002). The Marlowe-Crowne scale, consisting of 33 descriptions of highly desirable but rare and highly undesirable but common behaviors, measures respondents' tendency to present themselves in a positive light (Crowne & Marlowe, 1960). Initially conceptualized as unidimensional (Crowne & Marlowe, 1964), this scale has been suggested to be multidimensional, although there is no convergence on the number of factors (Barger, 2002; Loo & Loewen, 2004). For instance, Verardi et al. (2010) administered a shortened version of this scale in eight African countries and Switzerland, where they

distinguished achievement and international relationship; neither scale reached scalar invariance.

Millham (1974) and Ramanaiah, Schill, and Leung (1977) found a two-dimensional structure of the Marlowe-Crowne scale: enhancement (i.e., the tendency to attribute socially desirable characteristics to oneself) and denial (i.e., the tendency to deny undesirable characteristics). Such a distinction is in line with the two basic self-presentation motives: looking good and avoiding to look bad (Schütz, 1998). The two dimensions were found to have differential validity in predicting scales in the Minnesota Multiphasic Personality Inventory. According to Paulhus (1991) and Ventimiglia and MacDonald (2012), the Marlowe-Crowne scale taps mainly into impression management. Using the Balanced Inventory of Desirable Responding, an alternative two-dimensional structure of SDR was proposed by Paulhus (1984). He differentiated impression management (i.e., deliberate self-presentation to an audience) and self-deception (i.e., favorably biased but honestly held self-descriptions). The validity and utility of these two dimensions are not always supported in cross-cultural contexts (e.g., Helmes & Holden, 2003; Li & Bagger, 2006). Moreover, Paulhus and Reid (1991) reported that the distinction between enhancement and denial was more salient than between impression management and self-deception.

Individual- and Country-Level Variations of SDR

At individual level, education and socioeconomic status have been found to be negatively related to SDR (e.g., Uziel, 2010). Both males and females have the tendency to attribute socially desirable characteristics to themselves (Press & Townsley, 1998), with females often reporting higher SDR than males (Barger, 2002). There were no gender differences found in subdimensions of SDR such as enhancement and denial (Ramanaiah et al., 1977). Given the inconclusive findings regarding gender, we do not specify a directional hypothesis about gender differences, but explore these across cultural contexts.

Response styles in general have been found to be related to cultural values and personality traits (e.g., He, Bartram, Inceoglu, & van de Vijver, 2014; P. B. Smith, 2004). Lalwani, Shavitt, and Johnson (2006) reported that impression management, the main dimension tapped by the Marlowe-Crowne scale, was higher among collectivists than individualists. Schwartz et al. (1997) found a similar positive association between SDR and value types emphasizing social harmony in Finland and Israel. Trimble (1997) reported a positive association of SDR with intrinsic religiosity. Musek (2007) argued that SDR was positively related to the general factor of personality, a combination of the Big Five traits. At

country level, SDR was reported to be negatively associated with country affluence and individualism (T. P. Johnson & van de Vijver, 2003). Van Hemert et al. (2002) studied the Lie Scale from the Eysenck Personality Questionnaire (which is traditionally associated with SDR) in a cross-cultural meta-analysis, in which they confirmed the associations of the Lie Scale with affluence and individualism, and they reported a positive association with embeddedness measured with the Schwartz Value Survey. They also found a positive correlation with emotional stability and a negative one with extroversion at country level. If SDR indeed reflects valid individual and cultural differences, we expect that the aggregated values, beliefs and personality traits shared by individuals in each country are associated with SDR at country level.

The Present Study

It has been argued that it is difficult, if not impossible, to find scalar invariance of SDR across cultures (e.g, P. B. Smith, 2009), given that what is considered desirable varies from culture to culture. Comparing students from Singapore and United States, Li and Reb (2009) found weak support for the cross-cultural invariance of SDR in a multigroup confirmatory factor analysis; a similar conclusion was reached in a nine-country study (Verardi et al., 2010). With large-scale cross-cultural data, it is common to find nonequivalence and the underlying reasons are often unclear (Byrne & van de Vijver, 2010). It could be due to misspecification of the constructs in a few countries, accumulated small (even inconsequential) differences in parameters, or a combination of both. Some researchers argue that measurement invariance constraints in multigroup confirmatory factor analysis (i.e., invariance of loadings and intercepts) may be overly restrictive and that we need to allow for psychologically inconsequential variation in these parameters, such as done in Bayesian Structural Equation Modeling (Muthen & Asparouhov, 2012). We wanted to avoid these fit problems and did not want to use Bayesian Structural Equation Modeling given the lack of experience with its usage in empirical projects. Therefore, we resorted to an exploratory factor analysis approach (Costello & Osborne, 2005; Helmes & Holden, 2003) to study the factor structure of SDR with an adapted scale. The equivalence of the structure in different cultures was checked by means of calculations of Tucker's phi which is the congruence index of two sets of factor solutions (van de Vijver & Leung, 1997), followed by a differential item functioning (DIF) analysis to tease out items not suitable for cross-cultural comparisons.

We reasoned that one cause for the poor replicability of the factor structure of the Marlowe-Crowne scale is ambiguity in some items and potential inapplicability of some items in different cultures or with different populations. For instance, the original item "I never make a long trip without checking the safety of my car" does not apply to most people in less developed countries or to university students who do not own a car. Moreover, the wording in some items is redundant and outdated (e.g., "I don't find it particularly difficult to get along with loud mouthed, obnoxious people"), which poses challenges in precise translation to other languages. Given that some original items had limited discriminatory ability, various shortened versions of this scale have been proposed and validated (e.g., Reynolds, 1982; Strahan & Gerbasi, 1972). Similar to these previous studies, we used shortened and adapted items in the present study. Moreover, we aimed at maximizing cross-cultural comparability by adapting items.

Finally, to better understand the underlying mechanism of SDR, we studied the individual- and country-level correlates of SDR measured with this adapted scale in a multilevel design, taking into consideration data dependency at both levels. Specifically, we explored gender differences of SDR in cross-cultural contexts, and replicated and extended the study of the effects of country affluence, values, beliefs and personality traits on individual SDR.

Method

Participants

Participants were 3,471 university students with an age range of 17 to 35 in 20 countries. The mean age of these participants was 21.59 years (SD = 3.38). Thirty-one percent of the respondents were males. The sample size per country ranged from 95 to 389. The demographics are presented in Table 5.1.

Measures

The social desirability scale used in the present study was a shortened and simplified version of the Marlowe-Crowne scale. We selected items from the original scale with two criteria: items should not have ambiguous meaning and items should be appropriate in different cultural contexts. In addition, we simplified the original items to improve the cross-cultural comparability with translated versions. For instance, the original item "there has been times when I was quite jealous of the good fortune of others" was rephrased as "I am jealous of others with good fortune". There were 15 items in total, of which 9 items were worded as desirable attributes or behaviors (e.g., "I help others in trouble"), and 6 items were worded as

undesirable attributes or behaviors (e.g., "I gossip"). All items were formulated affirmatively to avoid artefacts from using negation (item keying). The wording comparison of the original items and the adapted items is presented in Table 5.2.

Table 5. 1 ***Demographics of the Participants***

Country	Sample Size	Mean Age (*SD*)	Percentage of Males	Language	Collection Mode
Bulgaria	194	20.53 (2.27)	23%	Bulgarian	1
China	374	21.12 (2.45)	48%	Chinese	1 and 2
France	389	19.05 (1.73)	20%	French	2
Germany	102	23.95 (3.01)	24%	German	1
Greece	167	25.60 (4.67)	32%	Greek	1
Indonesia	150	19.93 (1.08)	17%	English	2
Israel	98	27.70 (3.92)	36%	Hebrew	1
Italy	220	21.27 (0.63)	36%	Italian	2
Kenya	157	22.04 (2.63)	39%	English	2
Mexico	131	21.50 (3.94)	18%	Spanish	1
Netherlands	199	19.85 (2.34)	21%	Dutch	2
New Zealand	153	18.69 (1.91)	29%	English	1
Portugal	117	26.16 (4.73)	23%	Portuguese	1
Romania	193	22.53 (2.85)	17%	Romanian	1
Singapore	148	22.50 (2.09)	47%	English	1
South Africa	166	19.68 (1.60)	27%	English	1
Spain	106	20.77 (3.80)	36%	Spanish	1
Togo	201	22.48 (2.59)	50%	French	2
Turkey	95	23.49 (3.03)	22%	Turkish	1
United States	111	23.40 (3.48)	16%	English	1

Note. Collection mode: 1 = online; 2 = paper-and-pencil

To obtain sufficient psychometric details and allow more nuances in responses to this shortened scale, a 5-point Likert response format ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) was used. The original dichotomized (i.e., True versus False) and the Likert-scale responses to the Marlowe-Crowne scale have been shown to be comparable (Domínguez Espinosa, Salas Menotti, & Reyes-Lagunes, 2008) and support was found that a shortened version with the Likert-scale format did not result in substantial loss of information (Thunholm, 2001). Participants also provided demographic information such as their age and gender.

Country affluence was measured by the Human Development Index (HDI). The HDI is a composite measure of the average achievements in a country in three basic dimensions of human development: a long and healthy life, access to knowledge, and a high standard of living; it has a score range from 0 to 1, with a higher value indicating a higher level of

Table 5. 2 *Factor Structure of the 15-item Social Desirability Scale with the Pooled Sample*

Original Marlowe-Crowne Items	Items Adapted and Used in the Current Study	Enhancement	Denial
Before voting I thoroughly investigate the qualifications of all the candidates.	I think about my options before I make a choice	**.53**	-.06
I never hesitate to go out of my way to help someone in trouble	I help others in trouble	**.58**	-.16
It is sometimes hard for me to go on with my work if I am not encouraged	I continue with my work if I am motivated	**.53**	.04
On occasion I have had doubts about my ability to succeed in life	I have doubts about my ability to succeed in life	-.04	**.43**
I sometimes feel resentful when I don't get my way	I am satisfied when I get my way	**.50**	.34
I am always careful about my manner of dress	I am careful about my way of dressing	**.47**	.04
I like to gossip at times	I gossip	.22	**.59**
No matter who I'm talking to, I'm always a good listener	I am a good listener	**.55**	-.10
I sometimes try to get even rather that forgive and forget	I forgive others for their wrongdoings	**.39**	-.19
When I don't know something I don't at all mind admitting it	I admit when I do not know something	**.47**	-.18
At times I have really insisted on having things my own way	I do things my way	**.42**	.19
I would never think of letting someone else be punished for my wrongdoings	I let someone else be punished for my wrongdoings	-.17	**.56**
There have been times when I was quite jealous of the good fortune of others	I am jealous of others with good fortune	-.01	**.65**
I am sometimes irritated by people who ask favors of me	I am irritated by people who ask favors	-.06	**.53**
I have never deliberately said something that hurt someone's feelings	I say things that hurt others' feelings	-.07	**.57**

development (United Nations, 2012). Data for all 20 countries were available (denoted as N = 20).

Scores of country values, beliefs, and personality were extracted from various large-scale international studies. Scores of Hofstede's four cultural value dimensions (N = 17), measured from working adults, were taken from Hofstede (2009). Scores of Schwartz's seven culture-level value dimensions (N = 17), measured from teachers and students, were taken from Schwartz (2009a). The five belief dimensions in *Social Axioms* (N = 15) with students samples from Leung and Bond (2004) were also included. Scores of *Agreeableness, Conscientiousness, Emotion Stability, Extroversion,* and *Openness* (N = 13) were taken from the Big Five Inventory (Schmitt, Allik, McCrae, & Benet-Martínez, 2007). The respondents were mainly university students and in a few countries community-based samples.

Procedures

Administration procedures varied across countries. For countries where English is not the native language, the scale was translated, checked, and administered in the native language. In China, Mexico, and Singapore, the scale was administered on its own; in France, Togo, and Unites States, the scale was administered as part of a bigger study on education and career development; in all the other countries, the scale was administered as part of a bigger study on youth identity and well-being. In France, Indonesia, Italy, Kenya, the Netherlands, Togo and United States, data were collected with paper and pencil, whereas in the other countries an online survey was administered. In China, both data collection modes were used, with 274 respondents filled out the survey online and 100 with paper and pencil. The participation of all the students was voluntary.

Results

We describe the findings in three parts. First, we report the cross-cultural equivalence of the social desirability scale between the two data collection modes (within the Chinese sample). Second, we describe the cross-cultural equivalence and the item bias across countries. Finally, we report the multilevel analysis of SDR addressing gender and country characteristics.

Mode Effects

To account for possible administration bias caused by different data collection modes (Dwight & Feigelson, 2000), we first compared the factor structure and item means of the scale between the online survey and the paper-and-pencil sample within China. Principal component analysis with direct Oblimin rotation in either sample supported a two-factor solution: with all the positively worded items loading on the first factor (i.e., enhancement) and all the negatively worded items on the second factor (i.e., denial). Structural equivalence was evaluated with Tucker's phi (above .90 as acceptable and above .95 as excellent) (van de Vijver & Leung, 1997). The values of Tucker's phi for the two factors were 1.00 and 1.00 across modes, pointing to excellent structure invariance. Independent sample *t* tests on the 15 items between the collection modes showed no significant mean differences at $p < .01$. We then computed a DIF analysis, using a multiple regression approach. Each enhancement item was predicted with the total score on the enhancement factor, administration mode, and the interaction of the scale score and the administration mode, and each denial item was predicted with the total score on the denial factor, administration mode, and their interaction. The effect of the administration mode indicated uniform bias, and that of the interactions indicated non-uniform bias, both of which were evaluated by Cohen's f^2 when adding each predictor in multiple regressions (Cohen, 1988). Items with f^2 values larger than .15 (lower bound of medium effect size) were flagged as having non-negligible differential item functioning. In the current analysis, no items were found to display a medium or large effect size; the values of Cohen's f^2 ranged from 0 to .01. To summarize, the results indicated that the social desirability data collected in these two modes measure the same constructs and can be compared directly.

Structural Equivalence across Countries

Rather than computing all pairwise comparisons between countries, we chose for an approach in which we compared the factor structure of each country with the pooled sample (i.e., combining all participants and correcting for mean score differences on items by computing a weighted covariance matrix). Principal component analysis of the 15 items with direct Oblimin rotation was performed with the pooled sample (Table 5.2) and in each country. Two factors were extracted based on the scree plot, explaining 17% and 13% of the variance in the pooled sample respectively. The first five eigenvalues in the pooled solution were 2.56, 1.90, 1.25, .99, and .95, respectively. All the positively worded items loaded on the enhancement factor and all the negatively worded items loaded on the denial factor. The

explained variance of enhancement ranged from 15% to 21% and of denial from 11% to 15% in different countries. Similar to previous studies on the shortened Marlowe-Crowne scale (Verardi et al., 2010), the amount of variance explained in these factors was not high, possibly due to that response styles including SDR are broad, general tendencies, which can be difficult to capture wholly in questionnaire items. Structural equivalence was checked though comparing each country's factor solution with the pooled solution using Tucker's phi (Table 5.3).

Table 5. 3 ***Tucker' phi of the Factor Solutions between each Country and the Pooled Sample***

Country	Enhancement	Denial
Bulgaria	.50	.77
China	.98	.98
France	.98	.94
Germany	.96	.95
Greece	.91	.91
Indonesia	.85	.90
Israel	.88	.88
Italy	.98	.96
Kenya	.95	.94
Mexico	.97	.94
Netherlands	.96	.93
New Zealand	.96	.91
Portugal	.95	.89
Romania	.96	.87
Singapore	.92	.94
South Africa	.95	.94
Spain	.88	.93
Togo	.97	.95
Turkey	.93	.91
United States	.95	.92

On average the values of Tucker's phi for the two factors were .92 and .92. All the other countries showed acceptable structural invariance except Bulgaria, which might be due to the fact that we sampled students from the National Sports Academy who were not on an academic track as students majoring in social sciences in other countries. It was likely that the low convergence of the factor structure in Bulgaria resulted from the lack of familiarity in such survey tasks. We excluded Bulgaria in the following analyses.

Differential Item Functioning Analysis

Item bias across countries was examined using multiple regression analyses. We took South Africa as the reference group, because the adapted scale was first tested and validated in English in South Africa (van de Vijver & Meiring, 2011). Specifically, each of the 9 enhancement items was regressed on three blocks of variables: the deviance scale score of the enhancement factor, 18 country dummy variables, and 18 interactions between the country dummy variables and the deviance score of the enhancement factor. The same procedure was applied to the 6 denial items, which were predicted by the deviance score of the denial factor, the 18 country dummy variables, and the interactions between the two. To correct for the large number of predictors in the blocks, values of adjusted R^2 were used to calculate Cohen's f^2. Two denial items were found to have uniform bias with medium effect sizes (Table 5.4). On the item "I am irritated by people who ask favors", China, France, Germany, Indonesia, Israel, Italy, Mexico, New Zealand, Singapore, Togo, Turkey, and United States showed higher intercepts. On the item "I say things that hurt others' feelings", China, France, Portugal, Romania, and Togo showed higher intercepts whereas Greece had a lower intercept. These DIF effects could be due to different connotations of the words "irritated" and "hurt" in the various languages. These two items were excluded from the following analyses.

Table 5. 4 ***Effect Sizes in Regression Analyses: Uniform and Non-Uniform Bias Detection***

	Uniform Bias f^2	Non-Uniform Bias f^2
Enhancement Items		
I think about my options before I make a choice	.03	.01
I help others in trouble	.04	.00
I continue with my work if I am motivated	.04	.00
I am satisfied when I get my way	.10	.00
I am careful about my way of dressing	.10	.00
I am a good listener	.05	.00
I forgive others for their wrongdoings	.04	.01
I admit when I do not know something	.02	.00
I do things my way	.08	.00
Denial Items		
I have doubts about my ability to succeed in life	.07	.00
I gossip	.10	.00
I let someone else be punished for my wrongdoings	.09	.01
I am jealous of others with good fortune	.14	.00
I am irritated by people who ask favors	.16	.00
I say things that hurt others' feelings	.20	.00

The values of Cronbach's alpha of the final enhancement (9 items) and denial scale (4 items) were .62 and .54, respectively. The low reliability values were not unexpected; Beretvas, Meyers, and Leite (2002) in a reliability generalization study of the Marlowe-Crowne Social Desirability Scale reported that the estimated reliability of this scale was .53. Scale scores of the two subscales were calculated for each individual and aggregated to country level. The scores of the denial items were reverse-coded, thus a higher score on this subscale representing a higher tendency to deny negative self-descriptions. The two subscales were weakly correlated at individual level, $r(3274) = .09$, $p < .01$; their correlation at country level was nonsignificant, $r(17) = -.10$, $p = .69$, which could be due to the limited number of observations. The country scores of enhancement and denial are presented in Table 5.5.

Table 5. 5 ***Country Scores of Enhancement and Denial across 19 Countries***

Country	Enhancement	Denial
China	3.83	2.57
France	3.98	2.55
Germany	4.07	2.50
Greece	3.83	2.64
Indonesia	4.07	2.12
Israel	4.03	2.43
Italy	3.83	2.62
Kenya	4.06	3.01
Mexico	4.11	2.56
Netherlands	3.99	2.44
New Zealand	3.95	2.05
Portugal	4.14	2.88
Romania	4.29	2.64
Singapore	4.27	1.98
South Africa	4.16	2.80
Spain	4.09	2.38
Togo	4.00	3.22
Turkey	4.07	2.63
US	4.12	2.45

Multilevel Analysis

Before applying the multilevel analysis, we correlated the country-level scores of enhancement and denial with affluence, values, beliefs and personality traits. Due to the small sample sizes at country level and skewed distributions of some variables, we resorted to bootstrapping and the significance level of the correlations was determined in 1000 bootstrap samples (Table 5.6). Enhancement at country level was positively related to embeddedness and religiosity, and denial was negatively associated with HDI and positively associated with uncertainty avoidance, harmony, agreeableness, conscientiousness, and openness.

Table 5. 6 ***Country-Level Correlations of Enhancement and Denial with Affluence, Values, Beliefs, and Personality***

Country-Level Correlation	Enhancement	Denial
Human Development Index (*N* =19)	-.07	-.60*
Hofstede Values (*N* =17)		
Power Distance	.25	.15
Individualism	-.24	.02
Masculinity	-.15	.13
Uncertainty Avoidance	-.14	.60*
Schwartz Values (*N* =17)		
Harmony	-.18	.34*
Embeddedness	.37*	-.13
Hierarchy	.03	-.21
Mastery	-.34	.20
Affective Autonomy	-.20	-.12
Intellectual Autonomy	-.28	.17
Egalitarianism	-.19	.18
Social Axioms (*N* =15)		
Social Cynicism	.03	.23
Reward for Application	.30	-.26
Social Complexity	-.06	-.07
Fate Control	.11	-.09
Religiosity	.30*	-.22
Big Five Personality (BFI) (*N* =13)		
Agreeableness	-.08	.55*
Conscientiousness	.09	.60*
Emotion Stability	.38	-.03
Extroversion	.14	-.37
Openness	-.22	.32*

Note. *significance level established with 95% confidence intervals based on 1000 bootstrap samples

We tested the effects of gender and country characteristics on enhancement and denial in a multilevel design with HLM version 6 (Raudenbush & Bryk, 2002). The intraclass correlation coefficient was 9% for enhancement and 17% for denial, suggesting sufficient variations at country-level to conduct multilevel analyses (van de Vijver & Poortinga, 2002). In accordance with Enders and Tofighi (2007), we centered the country-level predictors (converted to the standardized *z* scores) on the grand mean. We entered one predictor per analysis. All multilevel analyses employed a random intercept and a fixed slope (Table 5.7).

Table 5. 7 ***Coefficients from Multilevel Analyses***

Predictor	Enhancement	Denial
Gender (male)	-.08**	.03**
Human Development Index (*N* =19)	-.01	-.15**
Hofstede Values (*N* =17)		
Power Distance	.03	.04
Individualism	-.03	.01
Masculinity	-.03	.04
Uncertainty Avoidance	-.02	.13**
Schwartz Values (*N* =17)		
Harmony	-.03	.08
Embeddedness	.05	-.03
Hierarchy	.00	-.05
Mastery	-.05	.05
Affective Autonomy	-.03	-.03
Intellectual Autonomy	-.04	.04
Egalitarianism	-.03	.04
Social Axioms (*N* =15)		
Social Cynicism	.00	.06
Reward for Application	.04	-.06
Social Complexity	-.01	-.02
Fate Control	.02	-.02
Religiosity	.04	-.05
Big Five Personality (BFI) (*N* =13)		
Agreeableness	.00	.12†
Conscientiousness	.03	.19†
Emotion Stability	.06	-.02
Extroversion	.02	-.10
Openness	-.02	.15

Note. *N* stands for the number of countries in the analysis.

†$p < .10$. **$p < .01$.

We first checked the effects of differences in data collection modes (online versus paper-and-pencil) on enhancement and denial, and found no significant differences. Compared with females, males scored lower on enhancement and higher on denial, indicating that there were differential effects of gender on the subdimensions of SDR. HDI was negatively associated with denial; its association with enhancement was nonsignificant yet in the expected (negative) direction.

Caution is needed in the interpretation of the country-level results, because the numbers of countries available for analyses with cultural values and personality traits were smaller (e.g., in most cases these data were not available for Kenya and Togo), which limited the cross-cultural variations that we could study. Albeit nonsignificant, power distance,

hierarchy, and emotion stability showed positive associations, and individualism and autonomy showed negative associations with enhancement, which well replicated the findings from van Hemert et al. (2002). Denial was predicted by agreeableness and conscientiousness. The combined evidence from the country-level correlations suggests that both enhancement and denial were related to cultural values and personality traits pertaining to "fitting in" and the two dimensions might be related to different aspects of "fitting in".

Discussion

We studied the factor structure, structural equivalence, and cross-country variations of a shortened and simplified Marlowe-Crowne Social Desirability Scale among university students in 20 countries. Our findings supported a two-dimensional structure of SDR, distinguishing the endorsement of positive self-description (enhancement) and the avoidance of negative self-description (denial). The structure was largely invariant across countries. There were gender differences in the two dimensions: enhancement was stronger among females and denial among males. A similar finding was reported by Sutton and Farrall (2005). It seems that there is a general difference in impression management in which females make more efforts to create a positive impression. There were more cross-country variations in denial than enhancement, and HDI was the most significant predictor for denial, which is in line with previous studies (T. P. Johnson & van de Vijver, 2003).

There has been much debate on the factor structure of SDR. Our study confirmed that SDR is a multidimensional construct. The distinction of enhancement and denial that we found in multiple countries is unlikely to be an artefact of item wordings, since all the items are formulated as affirmation of either positive or negative traits, emotions, and behaviors (Paulhus & Reid, 1991). The two aspects of SDR seem to be triggered by similar cultural mechanisms (i.e., fitting in). Given the small value of intraclass coefficients and the few significant predictors for enhancement, it seems that attributing positive traits to oneself is rather universal and is not much under cross-culturally differential control; however, denial has more cross-cultural variations, as people in countries low in affluence and high in agreeableness and conscientiousness seem to have a higher tendency to deny negative self-descriptions. These country-level correlates were similar to those found in a General Response Style (with SDR and extreme response style as positive indicators and acquiescent and midpoint response styles as negative indicators) (He, van de Vijver, et al., 2014; He & van de Vijver, 2013), suggesting that SDR, as part of a General Response Style factor, can be

interpreted as a means of response amplification motivated by “fitting in”.

Our study has a few limitations. We used data of SDR from various bigger projects with different administration modes, thus we did not have data on other constructs available in all countries that could be used to study the convergent and divergent validity of SDR at individual level. The university student sample may not be equally representative in each culture. In particular, access to higher education in less developed countries (e.g., Togo) is largely restricted to elites who do not necessarily reflect the values of the general population. In Bulgaria, students from non-academic track were sampled, which caused some incomparability with other countries. Future efforts should ensure the comparability of samples across cultures, and replicate the study with different conditions that may motivate or demotivate respondents to present themselves in a positive light (e.g., employment selection process). Yet we confirmed an equivalent structure of SDR within 19 countries and meaningful country-level correlates. Our study has important implications for cross-cultural research. First, we found that SDR measured by the adapted Marlowe-Crowne scale has a positive and a negative component that are weakly related to one another, but that do not show the same gender differences. Second, we find some systematic cross-cultural differences in enhancement and denial, which provides a piece to the puzzle as to whether SDR is nuisance or substance. We argue that SDR has at least some substantive meaning (McCrae & Costa, 1983), as the two dimensions, especially the denial dimension, are influenced by country affluence, cultural values and personality traits pertinent to “fitting in”. In such a case removing the effects of SDR can erroneously eliminate valid variations between individuals and cultures.

Chapter 6

Toward a Unification of Acquiescent, Extreme, and Midpoint Response Styles: A Multilevel Study

This chapter is based on
He, J., van de Vijver, F., J. R., Domínguez Espinosa, A. d. C., & Mui, P. H. C. (in press). Acquiescent, extreme, and midpoint response styles: A multilevel study. *International Journal of Cross-Cultural Management*. doi:10.1177/1470595814541424

Response styles are defined as a systematic tendency to use certain categories of the answering scale on some basis other than the target construct (Cronbach, 1950). They can impose validity threats in surveys, especially in cross-cultural studies where they can be sources of country differences in scores that are undesirable, yet difficult to control (e.g., Bachman & O'Malley, 1984). Self-report measures, prone to the influences of response styles, continue to be the most frequently used measures in cross-cultural management research (e.g., Donaldson & Grant-Vallone, 2002). Analyzing managerial data from 24 countries, van Emmerik et al. (2010) argued that cultural differences measured in the GLOBE leadership project and personality measures may represent culturally endorsed styles of responding. Therefore, investigating the nature of response styles can help cross-cultural management researchers establish more accurate measurements and derive more valid results.

We targeted the three most commonly studied response styles: acquiescent response style (ARS), extreme response style (ERS), and midpoint response style (MRS). ARS is defined as the tendency to agree rather than disagree to propositions in general; ERS is conceptualized as the tendency to endorse the most extreme response categories regardless of content; and MRS refers to the tendency to overuse the midpoint of a scale (Baumgartner & Steenkamp, 2001). There are two divergent views on response styles. The first is the traditional and still dominant perspective in which response styles are treated as measurement errors that should be avoided and eliminated as much as possible. An alternative view holds that response styles are a basic way of communicating about oneself, such as the tendency to amplify responses among Latin Americans and to moderate responses among East Asians. In the latter perspective, response styles are rooted in the values and personality of respondents and their cultures (Gibbons, Zellner, & Rudek, 1999; P. B. Smith, 2004, 2011); corrections for response styles may truncate valuable information embedded in the scores by removing reliable individual and cross-cultural differences (Fischer, 2004).

Response styles are particularly important in cross-cultural large-scale surveys. Analyzing data from the Teacher and Learner International Survey involving 23 countries, Vieluf, Kunter, and van de Vijver (2013) found that country differences in self-reported teacher self-efficacy could be largely explained by differences in response styles (notably ERS). Despite the prevalence of response styles and the more than 60 years that these have been studied, the psychological meaning of these styles is still unclear. The lack of progress may be a consequence of the emphasis on the need to remove or at least control for these response styles in the past. In addition, studies of response styles did not always show converging results, partly due to different operationalizations of these styles. For example,

agree responses were used in some studies as indicators of ARS and in some studies as part of ERS. We argue that integrating specific response styles to a General Response Style (GRS) may help create consistency in findings. Using both indirect and self-report measures of ARS, ERS, and MRS among members of five ethnic groups in the Netherlands, He and van de Vijver (2013) found a GRS with a positive loading of ERS and negative loadings of ARS and MRS at individual level, which showed sufficient measurement equivalence across groups. This GRS had a strong association with "desirable" personality traits (i.e., agreeableness, conscientiousness, extroversion, and openness).

In the present project, we gathered empirical evidence to examine whether this theoretically expected GRS can be found at both individual and country level, and how this GRS is correlated with relevant country-level characteristics in the first study, followed by a multilevel analysis in the second study. Our study has two novel aspects. Firstly, we set out to replicate the GRS at individual level and extend this to country level with data from multiple large-scale international surveys. Secondly, our study examined more surveys and country-level characteristics than previous studies. In the next section, we review the interrelatedness of the three response styles across cultures and postulate hypotheses in regard to the GRS with country-level characteristics.

Survey Response Styles Across Cultures

Interrelatedness of ARS, ERS, and MRS

At individual level, evidence suggests that ARS, ERS, and MRS are interrelated. MRS, the tendency to be evasive with a prevention focus, is negatively associated with ERS, the tendency to be outspoken with a promotion focus (van Vaerenbergh & Thomas, 2013). Some studies reported a weak positive association between ARS and ERS (e.g., van Herk et al., 2004) possibly due to the partly shared operationalizations of these two styles, in which agree responses are part of ARS and ERS scores. With data from various multicountry surveys, we expected to find the same patterning of a GRS at individual level (He & van de Vijver, 2013).

At country level, Smith (2011) reported a negative correlation between ERS and MRS, and a positive correlation between ARS and ERS; yet, the association between ARS and ERS may be overrated because of their partly shared operationalizations. It is reasonable to expect cross-level isomorphism (i.e., structural equivalence at various levels of aggregation) of their interrelatedness (van de Vijver & Poortinga, 2002). The dimension that goes from ERS to

MRS would represent the tendency to amplify or moderate responses (e.g., Minkov, 2009). We tested the hypothesis:

Hypothesis 1: At both individual and country level, there is a General Response Style (GRS) with a positive loading of ERS, and negative loadings of ARS and MRS.

Country-Level Correlates of Response Styles

Various social indicators, aggregated values, and personality traits show or can be expected to show relationships with response styles. These are reviewed below.

Socioeconomic development. Like many other psychological constructs, variations in response styles are associated with a cluster of affluence-related social indicators, such as the Human Development Index (HDI), literacy rate, democracy, and (absence of) corruption (e.g., van Dijk et al., 2009). The common denominator of these social indicators with relevance for response styles may be education and opportunities in life. More educated individuals may prefer to express their views in a more nuanced manner than individuals with less education, resulting in a lower level of amplifying response styles. We tested the following hypothesis:

Hypothesis 2: The GRS is negatively associated with socioeconomic development at country level.

Although our primary interest is in the GRS, we also tested this hypothesis (and the following hypotheses) for the other response style indexes. This means that the hypothesis implicitly predicts a negative association of socioeconomic development with ERS, and a positive association with ARS and MRS. We test the hypothesis about the GRS as well as the implied hypotheses, also for the following hypotheses.

Atheism and religious denomination. It has been argued that affinity to religion is linked to intolerance of ambiguity (Marshall & Lee, 1998). In countries where the majority of the population call themselves religious, people may tend to utilize higher level of response styles as a means of ambiguity reduction. Following this argument, we hypothesized:

Hypothesis 3: The GRS is negatively related to the percentage of atheists at country level.

Values. Previous studies found that response styles were positively related to a cluster of collectivistic values, including collectivism, embeddedness, and traditionalism (versus secularism) (P. B. Smith, 2004; van Dijk et al., 2009). Collectivistic cultures are characterized by an emphasis on group relationship and high context communication (Hofstede, 2001). Embeddedness is related to the levels that individuals identify themselves with the group, and the importance to maintain group traditions and restrain potentially disruptive action (Schwartz, 2009a). The transition from traditional to modern and postmodern societies is accompanied by a higher endorsement of secular values, which may link this transition to a decreased utilization of response styles (Inglehart & Welzel, 2005).

Hypothesis 4: The GRS is positively related to collectivistic values at country level.

Monumentalism (vs. flexumility) is characterized by national pride, acceptance of norms imposed by authority, and willingness to show superiority through interpersonal competition (Minkov, 2007). Smith (2011) found that ARS and ERS were positively correlated with monumentalism, whereas MRS showed a negative correlation. It seems that people in countries with high monumentalism orientation tend to amplify instead of moderate their responses. Based on Smith's work, we proposed:

Hypothesis 5: The GRS is positively associated with monumentalism at country level.

Personality characteristics. Equivalence of the five-factor model of personality at individual and country level has been established (McCrae et al., 2005a), therefore, we can use individual-level findings to formulate expectations about country-level associations between personality profiles and response styles. He and van de Vijver (2013) found that the individual-level GRS was positively associated with agreeableness, conscientiousness, extroversion, and openness, and negatively related to neuroticism. van Dijk et al. (2009) reported a positive correlation of ARS and ERS with extroversion. Harzing (2006) also found that extroversion was positively related to response styles in a consistent way. We hypothesized:

Hypothesis 6: The GRS at country level is positively associated with agreeableness, conscientiousness, extroversion, and openness, and negatively related to neuroticism.

The original conceptualization of social desirability as impression management shares important similarities with response styles (Paulhus, 1991), as both can be viewed as filters that are applied when generating item responses (e.g., He & van de Vijver, 2013). van Dijk et al. (2009) reported a positive correlation of ARS and ERS with the Eysenck Lie Scale, from which we speculated that these response styles would be akin to social desirability:

Hypothesis 7: The GRS at country level is positively associated with social desirability.

Study 1

In the first study, we set out to identify the GRS, using ARS, ERS, and MRS indexes constructed from eight cross-national surveys at both individual and country level, and examined the correlates of the GRS with various country-level characteristics.

Method

Data sources. We located eight multinational surveys with over 30 countries and with national representative samples. These surveys included European Value Survey (EVS, 2011), six waves of International Social Survey Programme (ISSP2003, 2004, 2005, 2006, 2007, and 2008 from www.issp.org), and World Value Survey (WVS, 2009). Detailed information of the surveys and uses of items for response style indexes are presented in Table 6.1. Participating countries in each survey varied in socioeconomic development levels (e.g., the mean level of Human Development Index HDI for participating countries in WVS ranged from .31 to .94 with a mean of .70).

Measures of ARS, ERS, and MRS. We computed indicators of ARS, ERS, and MRS, using a variety of items with Likert answer scales in each survey. Item content included attitudes towards personal values or beliefs (e.g., identity, leisure, and religiosity) and attitudes towards social issues (e.g., citizenship and governance). The average inter-item correlations among items for each response style index ranged from -.01 to .23, indicating sufficient content heterogeneity. Different items were used to compute each index. The non-overlapping items used in each response style index ensured data independency in testing the interrelatedness of response styles.

Table 6. 1 ***Overview of Surveys***

Survey	*N* of countries	*N* of respondents	*N* of items	*N* of Likert Points	Average Interitem Correlation
EVS: Social values	46 (24)	65911 (36892)			
ARS			14	5	.07
ERS			27	4	.18
MRS			13	5	.08
ISSP2003: National identity	35 (21)	45993 (29446)			
ARS			5	5	-.01
ERS			20	4	.19
MRS			24	5	.08
ISSP2004: Citizenship	38 (25)	52550 (35977)			
ARS			9	5	.05
ERS			8	7	.20
MRS			8	7	.18
ISSP2005: Work orientation	31 (19)	43440 (26842)			
ARS			18	5	.09
ERS			9	5	.07
MRS			9	5	.06
ISSP2006: Role of government	33 (21)	48641 (31319)			
ARS			8	5	.05
ERS			14	5	.03
MRS			14	5	.04
ISSP2007: Leisure, time and sports	34 (22)	49729 (31740)			
ARS			5	5	.06
ERS			11	5	.10
MRS			10	5	.12
ISSP2008: Religion	40 (27)	59986 (39855)			
ARS			9	5	.07
ERS			17	4	.17
MRS			12	5	.11
WVS: World values	48 (25)	66312 (35527)			
ARS			5	5	.15
ERS			16	4	.11
MRS			4	5	.23

Note. EVS = European Values Survey; ISSP = International Social Survey Programme; WVS = World Values Survey. Numbers in brackets indicated the sample sizes in Study 2.

ARS scores were derived from a set of items with five response options, ranging from 1 (*strongly agree*) to 5 (*strongly disagree*); the frequency of choosing 2 (*agree*) was taken as an ARS score (1 *strongly agree* was not used to avoid the confounding with ERS). ERS was derived from another set of items with various response options (e.g., *not at all* to *very much*, *not important* to *extremely important*) other than *agree-disagree*; the frequency of choosing the end points (e.g., 1 and 5 in a 5-point scale) was taken as an ERS score. MRS was constructed with a third set of items with various response options; the frequency of choosing the middle category (e.g., 3 in a 5-point scale) was taken as a MRS score. For each style, we first calculated the individual index for each respondent, and then averaged the index across members of the same country to obtain a country-level index.

Measures of country-level variables.

Socioeconomic development. The socioeconomic development of a country is measured here by the Human Development Index, the Gini index, and democratization. The *Human Development Index* (HDI) is a composite measure of the average achievements in a country in three basic dimensions of human development: a long and healthy life, access to knowledge, and a high standard of living (United Nations, 2012). Data were available for 194 countries (denoted here by N = 194). The *Gini index* measures the extent to which the distribution of income or consumption expenditure among individuals or households within an economy deviates from a perfectly equal distribution (N = 176) (The World Bank, 2011). *Index of Democratization*, the entitlement of ideologically and socially different groups to compete for political power, was obtained from the Polyarchy Index of Democracy (N = 184) (Vanhanen, 2007).

Atheism and religious denomination. *Religious denomination* was measured by the percentage of people who believe in God and the percentage of nominal adherents who celebrate traditional religious holidays (Wikipedia, 2007). The percentages of atheists, Muslims, and Christians were calculated for 220 countries.

Aggregated psychological profiles. *Individualism* scores were taken from Hofstede (2009) (N = 70). *Embeddedness* scores were taken from Schwartz's seven culture-level value dimensions (Schwartz, 2009a) (N = 75). The dimension scores of *Traditional Authority versus Secular-Rational Authority* were extracted from Inglehart's value dimensions (Inglehart, Basafiez, Diez-Medrano, Halman, & Luijkx, 2004). Higher scores on the dimension signified higher levels of secularism (N = 80). *Monumentalism* scores were taken from Minkov (2007). Higher scores on this dimension refer to countries that are proud and

self-consistent, whereas countries with lower scores exhibit the opposite tendency to show humility and flexibility ($N = 80$).

Scores of *Agreeableness, Conscientiousness, Extroversion, Openness, and Neuroticism* were taken from the Big Five Inventory ($N = 56$) (Schmitt et al., 2007). Scores of *Social Desirability* were taken from the Lie Scale of the Eysenck's Personality Questionnaire (van Hemert et al., 2002). The Lie Scale measures a tendency to fake good, which is a central aspect of social desirability ($N = 34$).

Finally, we also included citizen means of the five belief dimensions in *Social Axioms* from Leung and Bond (2004) for 39 countries, and the nine value dimensions from *GLOBE* (House, Hanges, Javidan, Dorfman, & Gupta, 2004) for 62 countries for exploratory purposes.

Results

We describe the results in two parts. Firstly, we report the analysis to identify the GRS, based on the ARS, ERS, and MRS indexes. Secondly, we correlated this GRS (i.e., the factor score) as well as each response style separately with country-level characteristics described below per domain: socioeconomic development, atheism and religious denomination, and aggregated values and personality.

Derivation of the GRS. The identification of the GRS was addressed in an exploratory factor analysis with the ARS, ERS, and MRS indexes per survey. At individual level, principal component analyses revealed a single factor in each survey; the explained variance ranged from 42% (WVS) to 46% (EVS). In all the cases, ERS loaded positively (loadings ranged from .19 to .82), and ARS (loadings ranged from -.04 to -.76) and MRS (loadings ranged from -.47 to -.82) loaded negatively on the factor.

At country level, a single factor was extracted in each survey with explained variance ranging from 47% (WVS) to 65% (ISSP2005). In all the cases, ERS loaded positively (loadings ranged from .77 to .91) and MRS loaded negatively (loadings ranged from -.75 to -.96) on the factor, whereas loadings of ARS varied across surveys (loadings ranged from -.56 to .55). The country-level GRS scores from each survey can be obtained from the first author.

We computed Tucker's phi coefficients to assess the similarity of the individual- and country-level factor solutions and found an average value of .90 (with a range from .79 to 1), which suggested an acceptable multilevel equivalence of the GRS (Van de Vijver & Poortinga, 2002). Hypothesis 1 was confirmed; the three response styles can be viewed as indicators of an underlying GRS with a positive loading of ERS and a negative loading of MRS (with ARS

in between). Yet, the analyses also clarified that the three styles are not interchangeable indicators and each has some uniqueness. We employed the factor score from each survey as the GRS at both individual- and country-level.

Correlation with country-level variables. We correlated the country-level GRS, ARS, ERS, and MRS in each survey with the country-level variables. Given the large number of correlation coefficients computed, we restrict the presentation to the mean correlations (Table 6.2), based on the Fisher transformation and back-transformation across the eight datasets. The significance level of these means was established in bootstrapping the means of the Fisher transformed correlations.

Socioeconomic development. We found that the GRS was negatively correlated with HDI and democratization, and positively correlated with the Gini index, suggesting that the GRS was higher in countries with lower socioeconomic development. Hypothesis 2 was supported.

Atheism and religious denomination. As hypothesized, GRS was negatively associated with the percentage of atheists and positively related to the percentage of Christians and Muslims, providing further support for the role of religion in response style use. Hypothesis 3 was supported.

Table 6. 2 ***Mean Correlations of the Response Styles with Country-Level Indicators***

Country-level indicator	GRS	ARS	ERS	MRS
Socioeconomic Development				
Human Development Index	-.50*	-.02	-.45*	.49*
Gini Index	.53*	.31*	.48*	-.50*
Democratization	-.39*	-.17*	-.31*	.41*
Religious Denomination				
Percentage Atheists	-.42*	-.25*	-.37*	.40*
Percentage Christians	.22*	.27*	.17*	-.20*
Percentage Muslims	.21*	-.12*	.24*	-.14*
Hofstede Values				
Power Distance	.18*	-.02	.19*	-.20*
Individualism	-.29*	-.12	-.26*	.36*
Masculinity	.05	-.03	.08	-.02
Uncertainty Avoidance	.14*	-.05	.18*	-.18*
Long Term Orientation	-.42*	-.23*	-.41*	.36*
Schwartz Values				
Harmony	-.23*	-.09*	-.26*	.20*
Embeddedness	.38*	.06	.37*	-.36*
Hierarchy	.24*	-.06	.25*	-.25*
Mastery	.14*	.08*	.15	-.17*
Affective Autonomy	-.34*	-.10	-.33*	.31*
Intellectual Autonomy	-.38*	-.14	-.36*	.34*

Table 6. 3 ***Mean Correlations of the Response Styles with Country-Level Indicators (cont.)***

Country-level indicator	GRS	ARS	ERS	MRS
Egalitarianism	-.05	.17	-.09	.07
Social Axioms				
Social Cynicism	.00	-.25*	.00	-.05
Reward for Application	.42*	.19*	.36*	-.40*
Social Complexity	-.09	-.06	-.09	.13*
Fate Control	.03	-.22*	.04	-.05
Religiosity	.27*	.18*	.29*	-.24*
Traditionalism versus Secularism	-.59*	-.38*	-.55*	.55*
Monumentalism	.41*	.24*	.40*	-.34*
GLOBE-should be				
Power Distance	-.01	.00	.02	.07
Uncertainty Avoidance	.40*	.04	.38*	-.43*
Institution Collectivism	.29*	.31*	.26*	-.33*
Ingroup Collectivism	.27*	.16*	.31*	-.22*
Gender Egalitarianism	-.05	.22*	.00	.07
Assertiveness	-.12	-.04	-.12	.07
Future Orientation	.35*	.16*	.37*	-.34*
Performance Orientation	.09	.38*	.12	-.10
Humane Orientation	-.02	-.24*	-.01	.07
Big Five Personality				
Extroversion	.06	.17*	.09	-.03
Agreeableness	.33*	.22*	.30*	-.23*
Conscientiousness	.35*	.14*	.36*	-.26*
Neuroticism	-.18*	-.14*	-.15*	.12
Openness	.21*	.22*	.25*	-.10
Eysenck Personality				
Psychoticism	-.23*	.00	-.23*	.10
Extroversion	.25*	.34*	.24*	-.17*
Neuroticism	-.10	.00	-.11	-.06
Social Desirability	.39*	-.16	.41*	-.35*

Note. The General Response Style factor (GRS), acquiescent response style (ARS), extreme response style (ERS), and midpoint response style (MRS). * $p < .05$ (significance level as determined in 1000 bootstrap samples).

Aggregated values. Individualism and secularism showed negative correlations with the GRS; Embeddedness showed a positive correlation with the GRS. Hypothesis 4 was supported. Monumentalism was positively associated with the GRS. Hypothesis 5 was supported. Though not hypothesized, there were consistent, positive correlations of the GRS with reward for application, GLOBE institution collectivism, ingroup collectivism, and future orientation, and negative correlations of the GRS with long term orientation and harmony.

Aggregated personality. The GRS was positively correlated with agreeableness, conscientiousness, and openness and negatively correlated with neuroticism, while its association with extroversion was nonsignificant but in the expected direction. Hypothesis 6

was largely supported. Finally, the GRS was positively related to social desirability, which provided support for Hypothesis 7.

Discussion

We examined the integration of ARS, ERS, and MRS from eight multicountry surveys and the associations of response styles with country-level characteristics in correlational analyses. The study demonstrated the existence and multilevel equivalence of the GRS, a combination of ERS, ARS, and MRS. Our findings suggest that it is meaningful to aggregate response styles in a GRS, in line with Cronbach's (1950) original definition, to a systematic tendency to use certain categories of the answering scale (in our conceptualization to amplify or to moderate expression).

Moreover, we found that the GRS is systematically associated with country characteristics; it is particularly salient in countries with a lower socioeconomic development, fewer atheists, and more religious people. The consistent correlations of the GRS with values and personality at country level suggest that response styles are part of national culture. People from cultures that value "fitting in" and ambiguity reduction more tend to use more GRS (i.e., more ERS and less MRS). Response styles can be viewed as communication filters that people use to express themselves. This communication filter is influenced by cultural characteristics. We tentatively summarize our findings by concluding that the GRS is positively associated with psychological variables that together make up two meaningful clusters: 1) "*fitting in*", as evidenced in the significant correlations with collectivism, embeddedness, traditionalism, agreeableness, and social desirability (e.g., P. B. Smith, 2004); and 2) *avoiding ambiguity*, as evidenced in significant correlations with percentages of religious people in a country, short-term orientation, reward for application, monumentalism, extroversion, and conscientiousness.

In addition to the GRS, we examined the associations of ARS, ERS, and MRS with country-level characteristics. ERS consistently showed the same patterning as the GRS, whereas MRS showed the opposite patterning. We did not find all expected associations for ARS. We expected and confirmed a positive loading of ERS and negative loadings of ARS and MRS on the GRS at individual level, but we found weak loadings and unstable positions of ARS at country level. It could well be that ARS, compared with ERS and MRS, is more domain specific (De Beuckelaer et al., 2010). Furthermore, ARS may have a slightly different meaning at individual and country level, which calls for further research.

Study 2

In the second study, we extended the investigation of the nomological network to a multilevel framework, in which we examined the influence of characteristics at individual level, country level, and their interactions on response styles.

Individual-Level: Predictors of Response Styles

The most frequently examined background characteristics at individual level with presumed relevance for response styles are age, gender, and education. Studies have shown that both ARS and ERS are more commonly used among elderly (e.g., Greenleaf, 1992b; Ross & Mirowsky, 1984) and the less educated (Marin et al., 1992; Rammstedt, Goldberg, & Borg, 2010), whereas MRS is more frequently used by the more educated (Sturgis, Roberts, & Smith, 2010). The correlates of ARS might be more difficult to evaluate in these studies given the possible confounding with ERS due to their common operationalizations, as argued above. Effect sizes of gender differences are inconsistent and small (van Vaerenbergh & Thomas, 2013). We used independent assessment of the response style indexes in examining the associations.

It may seem obvious to conclude that the GRS is positively related to age and negatively related to education. However, it should be noted that most studies were carried out in affluent countries. Interestingly, a study by Meisenberg and Williams (2008) employing secondary data from the WVS that involved a larger diversity of countries, found that the effects of age and education on response styles were not consistent across world regions. They reported a negative correlation of age and ERS in Africa and a positive correlation of education and ARS and ERS in South Asia. The findings cast doubt on the universality of the effects of age and education on response styles and suggest that country-level variables (i.e., socioeconomic development) could moderate the influence of response styles at individual level. Therefore, we explored the cross-level interaction.

Country-Level Predictors of Individual-Level Response Styles

Using the findings of the first study, we selected predictors among the country-level variables as most promising for a multilevel approach. We examined four predictors at the country-level: HDI, percentage of atheists, agreeableness, and monumentalism. These four predictors were chosen, because all of them had significant correlations with the four country-

level response style indexes, and each of them represented a cluster identified in the first study: socioeconomic development, religious denomination, "fitting in", and avoiding ambiguity.

Method

Data sources and measures. The same datasets from Study 1 were used in the second study. The sample sizes at each level for this study were shown in Table 6.1.

Three individual-level characteristics were examined in the multilevel analysis: age, education, and gender. Although the gender effect was less clear, we included this variable to explore its possible patterning across surveys. Age was recorded in years, education was measured by the highest degree achieved, and gender was dummy coded with 1 as male and 0 as female in all the eight datasets. The country-level predictors including HDI, percentage of atheists, agreeableness, and monumentalism were the same as in Study 1.

Analysis. We conducted multilevel analyses with HLM version 6 (Raudenbush & Bryk, 2002). Individual scores of the GRS, ARS, ERS, and MRS in each survey served as dependent variables. The intraclass correlation coefficients for the GRS ranged from 3% (WVS) to 11% (ISSP2004), for ARS from 5% (ISSP2005) to 15% (ISSP2003), for ERS from 7% (ISSP2006) to 16% (ISSP2008), and for MRS from 3% (ISSP2007) to 14% (WVS), suggesting that the threshold value of 5% proposed as the lower bound for conducting a multilevel analysis was reached in most cases (van de Vijver & Poortinga, 2002). In accordance with Enders and Tofighi (2007), we centered all individual-level predictors around their respective country means and the country-level predictors around the grand mean. All multilevel analyses employed a random intercept and fixed slope. For individual- and country-level predictors, we entered one predictor per analysis; for the cross-level interactions, we entered one individual- and one country-level predictor together with their interaction effect.

Results

We report the findings with respect to individual-, country-, and cross-level coefficients in the following section. A summary of the mean coefficients for each predictor and the cross-level interactions among the eight surveys is presented in Table 6.3. The significance level of parameters was established in a bootstrapping procedure.

Individual-level predictors. Age had a consistent, positive association with the GRS. Education had a consistent, negative effect on the GRS. The effect of gender on the GRS was nonsignificant.

Country-level predictors. HDI had a negative effect on the GRS. Percentage of atheists also had a negative effect on the GRS, whereas agreeableness and monumentalism showed positive effects. The predictions of country-level characteristics on individual-level response styles corresponded to the findings in the first study.

Cross-level interaction. Most interactions among individual-level predictors with country-level predictors were nonsignificant. We compared the correlations of the four response styles with age, education, and gender of low, medium, and high HDI countries in each survey, and did not find systematic differences in correlations. It can be concluded that the cross-level interactions were not salient.

Table 6. 4 ***Mean Regression Coefficients of Predictors in Multilevel Analyses***

Predictor	GRS	ARS	ERS	MRS
Age	.0055*	-.0001	.0005	-.0009*
Education	-.0404*	.0001	-.0039	.0058*
Gender	-.0365	.0099*	.0013	.0027
HDI	-1.7207*	.0008	-.3605*	.2386*
Percentage of Atheists	-.0068*	-.0006*	-.0011	.0010*
Agreeableness	.0255*	.0054*	.0080*	-.0039*
Monumentalism	.0016*	.0002*	.0004*	-.0002*
Age × HDI	.0097	.0019	.0010	-.0017*
Education × HDI	-.0400	-.0079	-.0188	.0038
Gender × HDI	.0024	.0335	-.0391	-.0329

Note. The General Response Style factor (GRS), acquiescent response style (ARS), extreme response style (ERS), and midpoint response style (MRS). Gender was coded as female as 0 and male as 1. HDI = Human Development Index. *$p < .05$ (significance level as determined in 1000 bootstrap samples).

Discussion

We studied the effects of both individual- and country-level predictors and their cross-level interactions on response styles in multilevel analyses. We found that age had a positive effect and education a negative effect on the GRS. The effect of gender was less clear. The mixed results are perhaps due to the characteristics of the items that we used to construct the indexes (e.g., different topics such as values, religion, work, and leisure; different levels of inter-item correlation which may confound the indexes and the substantive content measured in the items). Country-level HDI and the percentage of atheists showed negative effects, and agreeableness and monumentalism showed positive effects on the individual-level GRS,

supporting the substantive meaning and the cross-level equivalence of the meaning. The lack of significance in cross-level interactions points to the relatively stable effect of HDI (and other country-level characteristics) on the relationship of age, education, and gender with the GRS at individual level, even though we cannot rule out the possibility that there is an insufficient number of low HDI countries to test the interaction.

General Discussion

We set out to explore the patterning of a GRS extracted from ARS, ERS, and MRS at individual and country level. We confirmed that ARS, ERS, and MRS are interrelated response styles at both levels. The high level of multilevel equivalence suggests that the GRS has the same meaning at individual and country level. We found clusters of predictors of response styles at both levels, which points to the tendency to amplify or moderate responses. At country level, response style in general had the strongest association with 1) fitting in: low socioeconomic development, low percentage of atheists, collectivism, embeddedness, traditionalism, agreeableness, and 2) free from ambiguity: monumentalism, short-term orientation, reward for application, extroversion, and conscientiousness. At individual level, the response style factor was associated with age and education.

Interpretation of Response Styles

There has been much discussion on the interpretations of response styles: Are they measurement errors or do they have a more substantive meaning? The relationships at individual and country level suggest a meaningful patterning of response styles, probably related to communication styles which may have a deeper root in the socio-historical background of countries (Minkov & Blagoev, 2009). In our view, the dichotomy between measurement errors and substantive meaning is counterproductive. It is more adequate to see response styles as communication filters that impact all self-reports. The filter mainly works as the tendency to use specific parts of the response scales, either the middle or the extremes. The communication filter is related to impression management, related to aspects of "fitting in" and avoidance of ambiguity.

ARS is currently being measured in different ways in the literature. The common operationalization of ARS (i.e., using the endorsement of *agree* and *strongly agree* options) in most previous studies applied (e.g., Meisenberg & Williams, 2008) can create a spurious correlation with ERS. In this case, the interpretation of ARS is straightforward (as similar to

ERS). However, if ARS is only defined as agreement (using the endorsement of *agree* only, as in the present study), its meaning is not very clear. Our study suggests that ARS does not add much to the interpretation of the GRS. ERS and MRS are more defining characteristics.

Implication for Cross-Cultural Management Research and Practice

Accurate measurement is essential in advancing the field of cross-cultural management to deal with the challenge of response styles (Donaldson & Grant-Vallone, 2002). Our study on the integration and the psychological meaning of response styles shows that response styles are embedded in the values and personality of respondents and their cultures, therefore response styles cannot be easily "turned off". Instead, response styles can provide additional substantive information about individual and cross-cultural differences. For example, data of employees' personality traits collected from self-reports can be re-analyzed at individual level (i.e., constructions of response style indexes) to determine how they prefer to communicate. Aggregated response styles at a higher level would indicate the culture of an organization or society in large. In practice, methods such as score standardization to control for the effects of response styles in self-reports across cultures may not work adequately. Especially when the construct of interest is related to the clusters identified (i.e., fitting in, free from ambiguity), response styles can be considered a concomitant construct, related to the target construct, and score standardizations will erroneously remove variance in the target construct and influence the size of genuine cross-cultural differences.

Limitations and Future Directions

Despite the strengths and value of the current research, a few limitations and caveats should be mentioned; the post hoc construction of response style indexes may not be able to reflect the response styles precisely. We used data from a range of topics across surveys, and their associations with external variables converged, which to some extent ensured the validity of these indexes. Nevertheless, we are not sure that the substantive meaning measured in the construct and styles could always be adequately disentangled, notably in thematic surveys scales that tend to cover related domains. Moreover, response styles are not only associated with individual and cultural characteristics; the domain of questions being asked may also affect response style use (van Dijk et al., 2009), which was not considered in the present study. Future research may either develop direct measures of response styles or locate additional evidence to validate the indirect measures. With more reliable and valid measures, future efforts should study the domain dependency of response styles at both individual and culture level. To better utilize the GRS, a comparison of the effects of the GRS and other

methods to control for response styles (i.e., score standardization, anchoring vignettes) is also encouraged. In conclusion, the present research has tried to shed new light on the old, yet unresolved issue of the meaning and patterning of response styles. It is remarkable that after studying these styles for more than six decades the interpretation of response styles still looms large. We believe that progress in this field is contingent on attempts to better understand the mechanisms behind these styles and that cross-cultural research is important as it can provide variation in these styles which are beyond the horizon of monocultural research.

Chapter 7

Response Styles and Personality Traits: A Multilevel Analysis

This chapter is based on
He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology*, 45, 1028-1045. doi:10.1177/0022022114534773

This paper focuses on the association of response styles and personality traits in a cross-cultural context. Response styles are defined as respondents' systematic tendency to answer questionnaire items on some basis other than the specific item content (Paulhus, 1991). Response styles can create validity threats in cross-cultural studies, because observed score differences across cultures may not reflect true differences in the target construct, due to confounding response styles (e.g., van Herk et al., 2004). We investigate the shared and unique meaning of the four most studied response styles, namely acquiescent response style (ARS: the tendency to agree to a statement regardless of content), extreme response style (ERS: the tendency to overuse the end points of a scale), midpoint response style (MRS: the tendency to overuse the midpoint of a scale), and socially desirable responding (SDR: the tendency to answer questions in a way to make oneself look good). ARS, ERS, and MRS can be categorized as uniform response biases (i.e., the tendency to distort responses in a particular direction more or less regardless of the content of the stimulus), whereas SDR is more susceptible to the direction of the content (M. W. L. Cheung & Chan, 2002).

The consistency and stability of these response styles (e.g., Weijters et al., 2010) point to their trait-like properties and possibly even links to personality. Researchers have found associations among response styles and personality traits at both individual and country level. At individual level, ARS was associated with impulsiveness and extraversion (Couch & Keniston, 1960). ERS was positively related to intolerance of ambiguity, simplistic thinking, decisiveness, extroversion, and conscientiousness (Austin et al., 2006; Naemi et al., 2009). MRS was associated with evasiveness (Ayidiya & McClendon, 1990). SDR was associated with the general personality factor which consists of the combination of agreeableness, conscientiousness, extroversion, openness, and emotional stability (Bäckström, 2007; Schermer & MacDougall, 2013). However, a clear patterning of correlations between personality and response styles has not been established at country level, mainly because relatively few studies have been conducted. Harzing (2006) reported a positive association of extroversion with ERS and a negative one with MRS; McCrae et al. (2005a) reported a negative association of conscientiousness with ARS. P. B. Smith (2011) reported a negative association of openness with ARS and none of the Big Five personality traits related to ERS.

Findings on these reported associations could be challenged on two grounds. Firstly, the operationalizations of response styles were different in various studies. For instance, when constructed as the endorsement rate of the *agree* and *strongly agree* categories (e.g., 4 and 5 in a 5-point scale), ARS becomes a weak form of ERS (e.g., van Herk et al., 2004); when only the *agree* category is taken as ARS (e.g., 4 in a 5-point scale) to avoid confounding with ERS,

ARS is more related to MRS (e.g., He & van de Vijver, 2013). Secondly and more importantly, personality traits, often used as the validity measures of response styles, are typically measured with Likert scales, which themselves are likely to be tainted by the same response styles as are being investigated (Bentler et al., 1971).

The present study aims to address the two problems raised. To create consistency in the operationalizations of response styles, we integrate the four response styles in a General Response Style to achieve more robust estimates, in addition to examining each response style separately. We predict response styles with personality traits using the forced-choice format version of the Occupational Personality Questionnaire (OPQ32) provided by SHL, a talent measurement solutions company (SHL, 1999, 2006, 2009, 2013). SHL is a part of CEB (Corporate Executive Board), which provides talent management services for organizations all over the world. The major difference in the forced-choice format and the conventional Likert-scale personality instrument is that the former is robust to uniform response biases (i.e., ARS, ERS, and MRS), which allows us to measure country-level personality traits free from these uniform effects of response styles and to explain country variations in response styles that are found using other formats (e.g., Likert scales). In Study 1, we constructed country-level response style indexes and studied their integration and correlations with aggregated personality traits. Based on the country-level correlates identified, we applied a multilevel design in Study 2 to examine the unique variance explained by aggregated personality traits in individual-level response styles.

Study 1: Country-Level Analysis of Response Styles and Personality

We test the personality correlates of response styles and examine the convergence of results from personality traits measured with various formats. We focus on relationships of response styles with the forced-choice item format version of the OPQ32. We also use other personality instruments, which employ conventional Likert scales, to compare results of instruments that are more susceptible to influences of response styles, and to test the generalizability of our results.

Specific and General Response Style

It has been found that there is a predictable pattern of correlations among ARS, ERS, MRS, and SDR. ERS and MRS are negatively related to each other, reflecting the contrast of promotion and prevention focus (Cabooter, 2010). ERS and SDR are positively related, as

both are associated with desirable traits, such as extroversion and conscientiousness (Austin et al., 2006; Schermer & MacDougall, 2013). ARS, when operationalized as endorsing only the agree category, shows a negative correlation with ERS (Morren, Gelissen, & Vermunt, 2013). To integrate the full range of response styles, a General Response Style factor (GRS) can be extracted with ERS and SDR as positive indicators and ARS and MRS as negative indicators (He & van de Vijver, 2013). This General Response Style factor serves as a filter of questionnaire responses, by moderating or amplifying responses. In other words, the General Response Style factor has the function of a volume button that can play up or play down responses.

OPQ32

The OPQ32 measures 32 work-related personality traits defined as individual styles or preferences at work in three domains: relationships with people, feelings and emotions, and thinking style (see Table 7.1 for an overview) (SHL, 2009). It was developed with both a normative (Likert scales, OPQ32n) and a forced-choice version (OPQ32i, in which respondents choose one statement as *most like me* and one as *least like me* in blocks of four statements) (SHL, 1999, 2006). The OPQ32i was further evolved, resulting in the OPQ32r (SHL, 2013), which has 104 triplets selected from the original version. Responses to the triplets are scored as paired comparisons to which a multidimensional item response modeling model is applied. It can recover normative scale data from the forced-choice item format (SHL, 2006, 2013). Compared to the OPQ32n and OPQ32i, the OPQ32r measures the same constructs, demonstrates good psychometric properties and external validity, provides a good indication of the trait standing, and, like the OPQ32i, is robust to the effects of uniform response biases.

Through combining 26 out of the 32 personality scales in the OPQ32, the Big Five personality scores can be produced (SHL, 2006). Furthermore, the OPQ32 has been adapted for use in 30 languages, and the construct and scalar equivalence have been demonstrated across 39 countries (Bartram, 2013a, 2013b); thus, this country-level measure is proven to be suitable for cross-cultural comparative studies.

Table 7. 1 ***Definitions of the OPQ32 scales***

OPQ32	Definition
Domain 1. Relationships with People	
Persuasive	likes to change other people's views
Controlling	likes to be in charge
Outspoken	freely expresses opinions
Independent Minded	prefers to follow own approach
Outgoing	lively and animated in groups
Affiliative	likes to be around people
Socially Confident	feels comfortable when first meeting people
Modest	keeps quiet about personal success
Democratic	involves others in decision making
Caring	sympathetic and considerate towards others
Domain 2. Thinking Style	
Data Rational	enjoys analyzing statistical information
Evaluative	critically evaluates information
Behavioral	tries to understand motives and behavior
Conventional	prefers well established methods
Conceptual	enjoys discussing abstract concepts
Innovative	generates new ideas
Variety Seeking	likes changes to regular routine
Adaptable	changes behavior to suit the situation
Forward Thinking	takes a long-term view
Detail Conscious	likes to be methodical and organized
Conscientious	focuses on getting things finished
Rule Following	follows rules and regulations
Domain 3. Feelings and Emotions	
Relaxed	finds it easy to relax,
Worrying	worries about things going wrong
Tough Minded	not easily offended
Optimistic	expects things will turn out well
Trusting	sees others as reliable and honest
Emotionally Controlled	can conceal feelings from others
Vigorous	thrives on activity
Competitive	enjoys competitive activities
Achieving	likes to work to demanding goals and targets
Decisive	makes fast decisions

The Present Study

Based on what has been reviewed, we aimed to answer three research questions in the present study. Firstly, can a General Response Style factor, that was found previously at individual level (He & van de Vijver, 2013), be replicated at country level? We expect to find the same General Response Style factor with positive loadings of ERS and SDR and negative loadings of ARS and MRS at country level.

Secondly, what are the associations of response styles with expressions of personality traits that are free of uniform response biases? If response styles reflect culturally preferred

communication styles (e.g., P. B. Smith, 2004), we expect that the preference to dominate in relationships, characterized by high levels of being controlling and outspoken and low levels of being affiliative, modest, and democratic, would be positively related to response styles (in the direction of ERS and SDR versus ARS and MRS). In the feeling and emotions domain, we expect that being competitive and decisive would be associated with higher levels of response styles (Naemi et al., 2009). In the thinking domain, the preference to analyze statistical information and base inferences on facts and figures (i.e., data rational) to reduce ambiguity may boost the tendency to use response styles.

Thirdly, is there convergence in the associations of response styles with the Big Five personality traits measured with the forced-choice format (i.e., the OPQ32r) and other Likert-based inventories? If response styles only reflect scale usage, we would expect that a forced-choice questionnaire would show other correlations with measured personality traits than a Likert-based measure would. However, if response styles share trait variance with personality, we would expect more convergence between the two types of measures. Bartram (2013b) compared the 31 country-level Big Five scale means extracted from the OPQ32i with three other datasets: the Likert-based self-report NEO, the Likert-based other-report NEO, and the Likert-based self-report Big Five Inventory (BFI). He found moderate convergence among these four datasets.

Method

Data sources for ARS, ERS, and MRS. We located six waves of the International Social Survey Programme (ISSP2003, 2004, 2005, 2006, 2007, and 2008 from www.issp.org) with national representative samples from over 30 countries that had items suitable to construct ARS, ERS, and MRS indexes. Detailed information of the surveys and use of items for response style indexes is presented in Table 7.2.

We computed indicators of ARS, ERS, and MRS, using a variety of items that measured different constructs with Likert scales in each survey. The average inter-item correlations among items for each response style index ranged from -.01 to .20, indicating sufficient heterogeneity in item content (e.g., Greenleaf, 1992b). We used different items to compute each index, which ensured data independence and avoided statistical artifacts in testing the interrelatedness of response styles. ARS scores were derived from a set of items with five response options, ranging from 1 (*strongly agree*) to 5 (*strongly disagree*); the

Table 7. 2 ***Overview of Surveys***

Survey	*N* of countries	*N* of respondents	*N* of items	*N* of Likert Points	Average Interitem Correlation
ISSP2003: national identity	35	45993			
ARS			5	5	-.01
ERS			20	4	.19
MRS			24	5	.08
ISSP2004: citizenship	38	52550			
ARS			9	5	.05
ERS			8	7	.20
MRS			8	7	.18
ISSP2005: work orientation	31	43440			
ARS			18	5	.09
ERS			9	5	.07
MRS			9	5	.06
ISSP2006: role of goverment	33	48641			
ARS			8	5	.05
ERS			14	5	.03
MRS			14	5	.04
ISSP2007: leisure, time and sports	34	49729			
ARS			5	5	.06
ERS			11	5	.10
MRS			10	5	.12
ISSP2008: religion	40	59986			
ARS			9	5	.07
ERS			17	4	.17
MRS			12	5	.11

Note. ISSP = International Social Survey Programme. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style.

frequency of choosing 2 (*agree*) was taken as an ARS score (the endorsement of 1 *strongly agree* was not used to avoid the confounding with ERS).[1] Due to the limited number of Likert-scale items with the *strongly agree* to *strongly disagree* anchors available in the ISSP data sets, ARS indexes were constructed with smaller numbers of items compared with ERS and MRS. ERS was derived from another set of items with various response options (e.g., *not*

[1] We checked the correlations of the ARS indexes operationalized as the endorsement of the *agree* category (as in our study) and as the endorsement of both the *agree* and *strongly agree* categories (the conventional way) in the six ISSP data sets. At individual level, the correlations ranged from .55 to .77, with a mean of .65; at country level, the correlations ranged from .51 to .83, with a mean of .66. Thus, the ARS indexes in the current study correlate strongly with the ARS indexes that also include *strongly agree*.

at all to *very much*, *not important* to *extremely important*) other than *agree-disagree* scales; the frequency of choosing the end points (e.g., 1 and 5 in a 5-point scale) was taken as an ERS score. MRS was constructed with a third set of items with various response options; the frequency of choosing the middle category (e.g., 3 in a 5-point scale) was taken as an MRS score. For each style, we first calculated the individual index for each respondent, and then averaged the indexes across members of the same country to obtain a country-level index.

To get a global indication of country-level response styles, we replaced the missing values of the three response style indexes across the six surveys (20.45% missing) using EM imputation, given that the missing values were missing completely at random (Little's MCAR test: $\chi^2(120) = 137.90$, $p = .13$) (Little & Rubin, 2002). We treated the six indexes of each response style as scale items, and used the means across the six surveys as scores of ARS, ERS, and MRS. The values of Cronbach's alpha for the ARS, ERS, and MRS indexes were .85, .88, and .84, respectively.

Data sources for SDR. Scores on the country-level SDR were derived from an instrument-based meta-analysis (van Hemert, 2003) of the Marlowe-Crowne Social Desirability Scale from 42 countries (Domínguez Espinosa & van de Vijver, 2012). This scale was chosen because of its popularity over the years and prolific publications around the world. A total of 1052 publications from 1960 to 2011, which used the Marlowe-Crowne Social Desirability Scale or its various short forms, were reviewed. Data from 153 journal articles, 4 unpublished theses, and 1 technical report were retained after excluding studies with clinical, army, and inmate samples, and studies with "faking" or experimental instructions. Among all the studies, the scale was used in English (49.7%), French (10.1%), Spanish (8.2%), Chinese (6.9%), German (5.7%), Danish (3.8%), Portuguese (2.5%), Japanese, Swedish, Hebrew, Turkish and Arab/Iranian (1.9% each), Italian and Dutch (1.3% each), and Croatian, Greek, Norwegian, Indonesian, and Hindu (0.6% each). Means and standard deviations of the Marlowe-Crowne Social Desirability Scale were recorded for each study and aggregated to country level. We used the mean scores of SDR at country level in the present study (for more details of the meta-analysis and country rankings, refer to Domínguez Espinosa & van de Vijver, 2012).

Aggregated Personality Traits. The country-level sten scores (standardized scores ranging from 1 to 10, with a mean of 5.5 and a standard deviation of 2) of the 32 narrow personality traits as well as the Big Five personality traits from the OPQ32r in 39 countries were obtained from 113,480 working adults with over 20 languages in the data pool of SHL

(Bartram, 2013a). These data were collected as part of occupational assessments (i.e., for selection or development purposes in the workplace).

The country-level scale means of the Big Five personality traits from 31 countries with the self-report NEO were obtained from McCrae (2002). The respondents were college students and working adults. The country-level scale means of the Big Five personality traits from 50 countries with the other-report NEO were obtained from McCrae et al. (2005b). The respondents were college students who were instructed to rate someone they knew well.

The country-level scale means of the Big Five personality traits from 56 countries with the self-report BFI were obtained from Schmitt et al. (2007). The respondents were mainly college students and in a few countries community-based samples.

Results

We report the findings in three parts. We first examined the interrelatedness of the four response styles at country level and the extraction of a General Response Style factor; and then we correlated the response styles with the 32 personality scales from the OPQ32r; finally, we compared the correlation patterning of response styles with the four sets of Big Five personality data.

Table 7. 3 ***Correlations among the Response Style Indexes at Country Level***

	ARS	ERS	MRS	SDR
ARS	1			
ERS	.07	1		
MRS	-.21	-.83**	1	
SDR	.05	.34	-.60**	1

Note. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. SDR = Socially Desirable Responding.
**$p < .01$.

Interrelatedness among response styles. Table 7.3 presents the correlations among the four response styles at country level. ERS and SDR were positively related, and both were negatively related to MRS. The correlations of ARS with the other three response styles were nonsignificant. A one-factor solution was supported from the principal component analysis with the four response styles. This factor (explaining 54.76% of the variance) had positive loadings of ERS (loading .87) and SDR (.72), a negative loading of MRS (-.96) and a loading close to zero for ARS (-.05). This finding largely replicates the General Response Style (GRS) representing response moderation to amplification (He & van de Vijver, 2013). The only

difference with the previous study was that in the present study ARS had a negative loading closer to zero. We used the factor score as an index of GRS.

Table 7. 4 ***Zero-Order Correlation of Response Styles with the OPQ32r at Country-Level***

	GRS (*n* = 19)	ARS (*n* = 27)	ERS (*n* = 27)	MRS (*n* = 27)	SDR (*n* = 22)
Persuasive	.43	-.14	.55**	-.44*	.02
Controlling	.45	-.09	.65**	-.47*	-.02
Outspoken	.34	-.12	.48*	-.49**	.06
Independent Minded	.21	.06	-.23	.06	.48*
Outgoing	-.11	-.23	-.22	.23	-.21
Affiliative	-.48*	-.34	-.44*	.58**	-.51*
Socially Confident	-.37	-.08	-.10	.23	-.27
Modest	-.47*	.18	-.24	.29	-.02
Democratic	-.66**	-.11	-.64**	.58**	-.38
Caring	-.48*	.17	-.25	.32	-.28
Data Rational	.58**	-.18	.82**	-.62**	.17
Evaluative	.08	-.32	.24	-.06	-.27
Behavioural	-.08	-.22	.04	.28	-.35
Conventional	-.19	-.01	-.02	.10	-.12
Conceptual	-.04	.10	.15	.04	.05
Innovative	.23	.00	.50**	-.38*	-.08
Variety Seeking	.00	-.04	-.09	.06	-.15
Adaptable	-.41	.00	-.71**	.63**	.13
Forward Thinking	.02	-.01	.52**	-.33	-.01
Detail Conscious	-.30	-.15	.05	.07	-.19
Conscientious	-.22	-.16	.13	.01	-.28
Rule Following	-.11	-.16	.11	-.09	-.17
Relaxed	-.47*	-.28	-.32	.32	-.33
Worrying	.40	.17	.19	-.27	.16
Tough Minded	-.20	.23	-.05	.02	.01
Optimistic	-.32	-.06	-.15	.16	-.21
Trusting	-.44	-.02	-.32	.34	-.31
Emotionally Controlled	-.43	-.06	-.50**	.40*	.18
Vigorous	-.41	-.11	-.14	.29	-.31
Competitive	.64**	-.05	.74**	-.54**	.21
Achieving	.21	-.03	.52**	-.37	-.06
Decisive	.16	-.17	.29	-.33	-.12

Note. GRS = General Response Style. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. SDR = Socially Desirable Responding.
*$p < .05$. **$p < .01$.

Associations of response styles with the 32 personality scales. The response style indexes, including GRS, ARS, ERS, MRS, and SDR, were correlated with the 32 scales of personality traits from the OPQ32r (Table 7.4). GRS was positively related to Competitive

and Data Rational, and negatively related to Affiliative, Modest, Democratic, Caring, and Relaxed. Although not statistically significant, Persuasive, Controlling, and Worrying showed positive associations, and Adaptability, Trusting, Emotionally Controlled, and Vigorous showed negative associations with GRS. The contrast in correlational patterns between ERS and MRS was most salient; ERS had positive associations with Persuasive, Controlling, Outspoken, and Innovative, and negative associations with Affiliative, Adaptable, Democratic, and Emotionally Controlled, whereas MRS showed an opposite pattern. ARS was not significantly associated with these personality traits. SDR was positively related to Independent Minded and negatively related to Affliative.

Associations of response styles with the Big Five personality traits. We correlated the response style indexes with the Big Five personality traits measured by the OPQ32r, self-report NEO, other-report NEO, and self-report BFI (Table 7.5). GRS was strongly correlated with the OPQ-based agreeableness (negative) and the NEO-based self-report conscientiousness (positive). Emotional stability and extroversion from all the four sets of data were unrelated to any of the response styles. However, the main finding of the analysis was the weak convergence of correlations across instruments. Across all traits, correlations between the correlations shown in Table 7.5 across instruments were only significant for the two NEO measures, $r(25) = .80$, $p < .001$. Notably the absence of any relationship between the two Likert-based self-report measures is remarkable, even if a lack of significant correlations is a recurrent problem in this field. Moreover, we could not replicate the association of extroversion with ERS and MRS that was found by Harzing (2006), probably due to the different personality measures used (in Harzing's study extroversion was from the Eysenck Personality Questionnaire that stresses expressiveness and liveliness). Contrary to McCrae et al. (2005a), we found a positive association of ARS with conscientiousness, which could be explained by the very different operationalization of ARS in their study (i.e., the sum of the raw scores of the 240 NEO personality items as ARS). The failure to replicate the negative association of ARS with openness as in Smith (2011) may be attributed to a similar cause (ARS there was calculated using twice the number of scale point "5" responses, plus the number of scale point "4" responses in a 5-point scale). Our findings suggest that response styles are instrument specific; yet, we cannot rule out the impact of sample incomparability as all measures are based on convenience sampling.

To summarize, we found some significant personality correlates of response styles, and we did not find strong convergence of the associations between response styles and Big Five personality traits. However, it seems that even if we use a response scale that cannot

reflect ARS, ERS, or MRS, the Big Five personality scales were still related to response style indexes, which strongly suggests that these indexes have a substantive meaning.

Table 7. 5 ***Correlations of Response Styles with Big Five Personality Traits***

		GRS	ARS	ERS	MRS	SDR
Agreeableness	OPQ32	-.71**	.01	-.54**	.54**	-.49*
	Self-report NEO	-.38	.18	-.22	.51*	-.03
	Other-report NEO	-.14	.12	-.20	.18	-.02
	BFI	.40	.28	.50*	-.32	-.03
Conscientiousness	OPQ32	-.26	-.14	.17	.01	-.27
	Self-report NEO	.69**	.53*	.69**	-.50*	.44
	Other-report NEO	.33	.15	.23	-.39	.21
	BFI	.32	.18	.42*	-.24	-.05
Emotional Stability	OPQ32	-.45	-.14	-.23	.27	-.26
	Self-report NEO	-.20	.00	.07	.25	-.06
	Other-report NEO	-.16	.38	-.14	.20	-.05
	BFI	.19	.35	.22	-.09	.25
Extroversion	OPQ32	-.02	-.23	.16	.01	-.31
	Self-report NEO	-.02	.03	.29	.08	-.42
	Other-report NEO	-.16	.09	.14	.19	-.58**
	BFI	-.03	.22	.04	-.03	.04
Openness	OPQ32	.15	-.02	.23	-.13	-.07
	Self-report NEO	-.22	.15	-.36	.29	-.09
	Other-report NEO	-.31	.33	-.45*	.33	-.01
	BFI	.52*	.40	.56**	-.38	-.05

Note. *N* of countries = 19, 27, 27, 27, and 22 for GRS, ARS, ERS, MRS, and SDR in OPQ32;
N of countries = 15, 20, 20, 20, and 18 for GRS, ARS, ERS, MRS, and SDR in self-report NEO-PI-R;
N of countries = 15, 21, 21, 21, and 19 for GRS, ARS, ERS, MRS, and SDR in other-report NEO-PI-R;
N of countries = 16, 23, 23, 23, and 20 for GRS, ARS, ERS, MRS, and SDR in BFI.
GRS = General Response Style. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. SDR = Socially Desirable Responding.
*$p < .05$. **$p < .01$.

Discussion

We carried out country-level correlational analyses with response styles extracted from national representative respondents from multinational surveys and personality traits from the OPQ32r and other Likert-based inventories. We found a General Response Style factor (i.e., the shared meaning of ERS and SDR versus ARS and MRS) that reflected the tendency to amplify or to moderate responses. The significant correlations of some but not all personality traits measured with the forced-choice OPQ32r supported our expectation that response styles have some substantive meaning, given that the uniform response style bias cannot influence this instrument and response styles such as ARS, ERS, and MRS cannot be computed on the basis of the force-choice instrument. In other words, scores on the OPQ32r

cannot reflect scale usage and should have some substantive meaning if there would be significant correlations between scores on this personality inventory and response style indexes. Specifically, we found support that response styles at country level reflect aspects of culture (P. B. Smith, 2004), as evidenced in (1) dominating versus submitting to interpersonal relationship, (2) competitive versus modest and democratic feelings and emotions, and (3) data rational thinking. Therefore, response styles can be viewed as culturally moderated communication filters that amplify or moderate responses. Response styles cannot be dismissed as only reflecting scale usage; instead, it appears that the way in which individuals use scales is telling of their personality. These effects are even more pronounced when scores are aggregated at country level.

Study 2: A Multilevel Analysis of Response Styles

Building upon Study 1, we studied the psychological meaning of individual-level ARS, ERS, and MRS (individual-level data on SDR were not available, thus not included) in a multilevel design with country-level OPQ32r scale scores.

Individual (Demographic) Correlates of Response Styles

The effects of age, gender, and education on response styles have been frequently examined. Some studies reported that both ARS and ERS were more commonly used among elderly (Greenleaf, 1992a; Meisenberg & Williams, 2008) and the less educated (Marin et al., 1992), whereas MRS was more frequently used by the more educated (Sturgis et al., 2010). Yet some studies reported the opposite effects; for example, Austin et al. (2006) found a negative effect of age on ERS. Effects of gender were generally inconsistent and small (van Vaerenbergh & Thomas, 2013). A possible explanation for such conflicting results could be the sampling scheme in different studies (using either convenience or probability samples). The national representative samples from the International Social Survey Programme, as used in the current study, are expected to have adequate variations on these background variables and findings based on such samples may afford better ground for generalization.

Country-Level Correlates of Response Styles

We investigate the socioeconomic development, the overarching cultural value (i.e., individualism), and aggregated personality traits at country level that were found to be relevant for response styles. It has been reported that the socioeconomic development of a

country has a strong relationship with response styles (van Hemert et al., 2002). It may be related to individual opportunities in life; people living in countries with lower levels of socioeconomic development may need to use whatever accessible resources and strategies to survive and thrive, resulting in more use of amplifying response styles.

Individualism (versus collectivism), the most frequently studied cultural value dimension (Hofstede, 1980, 2001), was found to be associated with response styles in many studies. Yet findings were not unequivocal. Harzing (2006) reported a negative association of individualism with ARS and MRS, and no association with ERS. De Jong, Steenkamp, Fox, and Baumgartner (2008) reported a positive association of individualism with ERS. Van Herk et al. (2004) reported a negative association of individualism with ERS and ARS. Again, sampling frames could be a confounding factor. We explore the effects of individualism on response styles from our multiple national representative samples.

Drawing from the first study, we selected four personality traits at country level: Affiliative, Democratic, Data Rational, and Competitive. These four personality traits were chosen, because they showed the strongest and most consistent correlations with GRS at country level.

Method

Data sources of individual-level response styles. We used the same ISSP data sets as in Study 1. We calculated the individual level ARS, ERS, and MRS in each of the six ISSP data sets. We factor analyzed the three response styles at individual level in each set, and found a one-factor solution (with explained variances ranging from 48.51% to 65.10%) with a positive loading of ERS (loadings ranged from .19 to .82 with a mean of .70) and negative loadings of ARS (from -.04 to -.76 with a mean of -.35) and MRS (from -.72 to -.82 with a mean of -.78) for all the six data sets. Pair-wise comparisons of the structural equivalence of the factor solutions in the six datasets were carried out by means of the calculation of Tucker's phi (van de Vijver & Leung, 1997). We found an average value of .90, indicating acceptable structural equivalence. The factor scores were taken as the GRS scores in this study. In addition to these individual-level response styles, we used five dummy variables to gauge ISSP survey-specific effects.

Individual-level predictors. For each respondent in all the six ISSP surveys, data on age (in years), gender (male as 1 and female as 0), and education measured as the highest degree achieved from 1 (*no formal qualification*) to 5 (*university completed*) were available.

Country-level predictors. The *Human Development Index* (HDI) was used to indicate the socioeconomic development of a country (United Nations, 2010). *Individualism* scores for 70 countries were taken from Hofstede (2009). The sten scores of Affiliative, Democratic, Data Rational, and Competitive from the OPQ32r were the same as in Study 1.

Analysis

We conducted the analyses with HLM version 6 (Raudenbush & Bryk, 2002). In total, data covered 184,433 respondents in 26 countries. The values of the intraclass correlation coefficients were 13% for GRS, 6% for ARS, 15% for ERS, and 10% for MRS, suggesting enough variation at country level to conduct multilevel analysis (van de Vijver & Poortinga, 2002).

In accordance with Enders and Tofighi (2007), we standardized all predictors to *z* scores, and centered all individual-level predictors around their respective country means and the country-level predictors around the grand mean, except for dummy variables at individual level to distinguish survey differences which were neither standardized nor centered.

In line with the model building proposed by Nezlek (2008), we used the forward-stepping procedure to add predictors, first with individual-level variables and then country-level variables for each model. To be specific, we first checked if there were mean differences across surveys by adding the dummy variables, and then we added age, gender, and education at individual level. Next, we added HDI, then individualism, and finally each of the chosen personality traits from the OPQ32r. As Bartram (2013b) reported that country-level OPQ32 scales were strongly associated with HDI and individualism, adding country-level predictors in such a sequence would allow us to examine the unique variance explained by the aggregated personality traits. We evaluated the significance of the change of Chi-square statistics after adding each predictor, from which we determined whether to retain the predictor in the model. All multilevel analyses employed random intercepts and fixed slopes.

Results

Table 7.6 presents the model building for GRS, ARS, ERS, and MRS. We excluded country-level variables that resulted in an increase in Chi-square statistics for each response style index.

Table 7. 6 ***Model Building in Multilevel Analysis***

	χ^2	df	Δχ2	Δdf	
(a) Model comparison for GRS					
(1) HDI	8615	24	-	-	-
(2) HDI + IND	6776	23	1839	1	(2)-(1)
(3) HDI + IND + ***Affiliative***	6977	22	-201	1	(3)-(2)
(4) HDI + IND + ***Democratic***	6937	22	-161	1	(4)-(2)
(5) HDI + IND + Data Rational	3935	22	2841	1	(5)-(2)
(6) HDI + IND + Data Rational + Competitive	3539	21	396	1	(6)-(5)
(b) Model comparison for ARS					
(1) HDI	6058	24	-	-	-
(2) HDI + ***Individualism***	6079	23	-21	1	(2)-(1)
(3) HDI + Affiliative	4669	23	1389	1	(3)-(1)
(4) HDI + Affiliative + ***Democratic***	4675	22	-6	1	(4)-(3)
(5) HDI + Affiliative + ***Data Rational***	4675	22	-6	1	(5)-(3)
(6) HDI + Affiliative + ***Competitive***	4682	22	-13	1	(6)-(3)
(c) Model comparison for ERS					
(1) HDI	9911	24	-	-	-
(2) HDI + Individualism	8362	23	1549	1	(2)-(1)
(3) HDI + Individualism + ***Affiliative***	8544	22	-182	1	(3)-(2)
(4) HDI + Individualism+ Democratic	7448	22	914	1	(4)-(2)
(5) HDI + Individualism + Democratic + Data Rational	3757	21	3691	1	(5)-(4)
(6) HDI + Individualism+ Data Rational + Democratic + Competitive	2962	21	795	1	(6)-(5)
(d) Model comparison for MRS					
(1)HDI	5258	24	-	-	-
(2)HDI + Individualism	3568	23	1690	1	(2)-(1)
(3)HDI + Individualism + ***Affiliative***	3678	22	-110	1	(3)-(2)
(4)HDI + Individualism + ***Democratic***	3630	22	-62	1	(4)-(2)
(5)HDI + Individualism + Data Rational	2633	22	935	1	(5)-(4)
(6)HDI + Individualism + Data Rational + Competitive	2584	21	49	1	(6)-(5)

Note. Italicized variables were excluded in the model due to the increase of the Chi-square values when they were added. GRS = General Response Style. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. SDR = Socially Desirable Responding. HDI = Human Development Index.

Table 7.7 presents the regression coefficients for each response style. At individual level, means of indexes varied across surveys (as indicated in the significant regression coefficients in the survey-specific dummy variables), which pointed to domain specific effects of response styles (van Dijk et al., 2009). Age was positively related to GRS and ERS, and negatively related to ARS and MRS. Males scored lower on GRS, and higher on ARS and MRS. Education had a negative effect on GRS and ERS, and a positive effect on MRS. However, all these effects were small.

Table 7. 7 ***Regression Coefficients in Multilevel Analysis***

	GRS	ARS	ERS	MRS
Individual-Level Predictors				
Age	.0960**	-.0027**	.0083**	-.0122**
Gender (male)	-.0111**	.0044**	.0017	.0016**
Education	-.0442**	-.0008	-.0047**	.0057**
ISSP 2004	-.0137†	.4132**	.1100**	-.0888**
ISSP 2005	.0207**	-.0151**	-.1290**	-.0551**
ISSP 2006	-.0802**	.0248**	-.1808**	.0543**
ISSP 2007	-.0856**	.0454**	.0037	.0592**
ISSP 2008	.0419**	-.0388**	.1506**	.0830**
Country-Level Predictors				
HDI	-.1870**	.0442**	-.0303**	.0227**
Individualism	-.0895†	-	-.0162*	.0143**
Affiliative	-	-.0387**	-	-
Data Rational	.1232*	-	.0235*	-.0132*
Democratic	-	-	.0024	-
Competitive	.0473	-	.0219†	-.0025

Note. GRS = General Response Style. ARS = Acquiescent Response Style. ERS = Extreme Response Style. MRS = Midpoint Response Style. HDI = Human Development Index.
†$p < .10$. *$p < .05$. **$p < .01$.

At country level, GRS was characterized by low HDI, low Individualism, high Data Rational, and high Competitive; the same patterning was found for ERS, and the opposite patterning for MRS. ARS was predicted only by high HDI and low Affiliative.

Discussion

Using national representative samples from 26 countries, we studied the individual-level response styles in associations with age, gender, and education at individual level, and socioeconomic development, individualism, and aggregated personality traits at country level. The regression coefficients for GRS at individual level were consistent, and they were stronger than those of specific response styles, indicating that the use of a General Response Style results in more robust estimates. The effects of age and education were largely in line with Meisenberg and Williams (2008); older and less educated respondents tend to use amplifying response styles more, whereas younger and more educated respondents tend to use moderating response styles more. It seems that being from a younger generation with higher education is associated with a preference for more nuanced views, which may work against the use of ERS and may be more in line with preferences of middle categories of response scales. Gender differences were much smaller than those of age and education; therefore, gender is a less salient predictor for response styles.

At country level, HDI and individualism were negatively associated with ERS and positively associated with MRS. In our view, this patterning suggests that individuals in less affluent and more collectivistic countries show a stronger tendency to stress their conformity and to amplify what the members of their group are supposed to respond. Aggregated personality traits, although closely linked with cultural values (Bartram, 2013b; Hofstede & McCrae, 2004), could explain additional variance in response styles, which furthered our understanding of the mechanism underlying response styles. The positive effects of Competitive on GRS and ERS and the opposite effect on MRS demonstrate that response amplification is more prevalent in countries where people enjoy competitive activities more. The effects of the Data Rational scale provide evidence for the rational thinking based amplifying response style. It seems counterintuitive that GRS is positively associated with data rational thinking, as data rational thinking is expected to be positively related to education level at individual level and affluence at country level. Enjoying the analysis of statistical information and basing decision on facts and figures may be positively associated with education; however, the unwillingness to deal with opinions and feelings points to rigidity and inflexibility, which could lead to the endorsement of ERS rather than ARS and MRS. To conclude, the multilevel analysis of the large set of response style with predictors from both levels presents a complex picture of the response style constructs.

General Discussion

We studied the psychological meaning of response styles in a correlational and a multilevel study with the OPQ32r and other country characteristics. We found that response styles are interrelated, thus a General Response Style can be extracted with positive loadings of ERS and SDR (when available), and negative loadings of ARS and MRS, at both individual and country level. The general as well as specific response styles are systematically related to socioeconomic development, individualism, and aggregated personality traits (in particular Competitive and Data Rational) from the OPQ32r at country level. At individual level, response styles are systematically related to age and education.

Interpretation of Response Styles

There has been an ongoing debate on the nature of response styles. On the one hand, response styles are viewed as mere measurement artifacts; they may pose serious challenges

to the validity and comparability of scores (Clarke III, 2001) or have negligible effects (Ones et al., 1996). On the other hand, response styles are a promising candidate in characterizing aspects of national cultures (P. B. Smith, 2009), as they may represent preferred ways of communication that are embedded in individual and cultural values and personality traits (Harzing, 2006; P. B. Smith, 2004, 2011). Our findings provide some evidence for the substance interpretation of response styles. The personality traits measured by the forced-choice format OPQ32r provided the unique opportunity for response styles to manifest themselves as systematic effects on scores other than uniform response biases that prevail in the analysis of Likert-based scales. At country level, there are significant positive associations of GRS (i.e., ERS and SDR versus ARS and MRS) with Competitive and Data Rational, and negative associations with Affiliative, Democratic, and Modest. Moreover, controlling for the effects of socioeconomic development and individualism, we still see unique effects of Competitive and Data Rational on individual-level response styles, therefore it is reasonable to conclude that response styles are more about substance than nuisance (e.g., Schwartz et al., 1997).

Since we found significant correlations of response styles with traits measured by a personality instrument that cannot display ARS, ERS, or MRS, response styles seem to be integrated in personality traits or at least their expression. Thus, response styles may not be a red herring as suggested by Ones et al. (1996). In practice, researchers should be cautious in applying any measure to control for response styles (Fischer, 2004). If we were to correct for such response styles, we would eliminate valid personality variance. We conclude that GRS, the integration of the four styles, can be interpreted as a filter that influences all self-reports, yet the extent of the influence may vary across constructs. Response styles share trait variance with constructs related to dominance, competitiveness, and data rational thinking; therefore, response styles are better not be corrected for in the measurement of these constructs.

Convergence and Divergence of Personality Instruments

We set out to study response styles with a personality instrument that is robust to the effects of uniform response biases, and found different levels of convergence with other personality instruments. In predicting response styles with the 32 individual personality traits of the OPQ32r, the significant personality trait correlates were in line with expectation, yet the aggregated Big Five personality traits showed a less consistent pattern. The lack of convergence may be due to sample differences (e.g., small country-level sample sizes, incomparable samples from convenience sampling), instrument characteristics (e.g., different

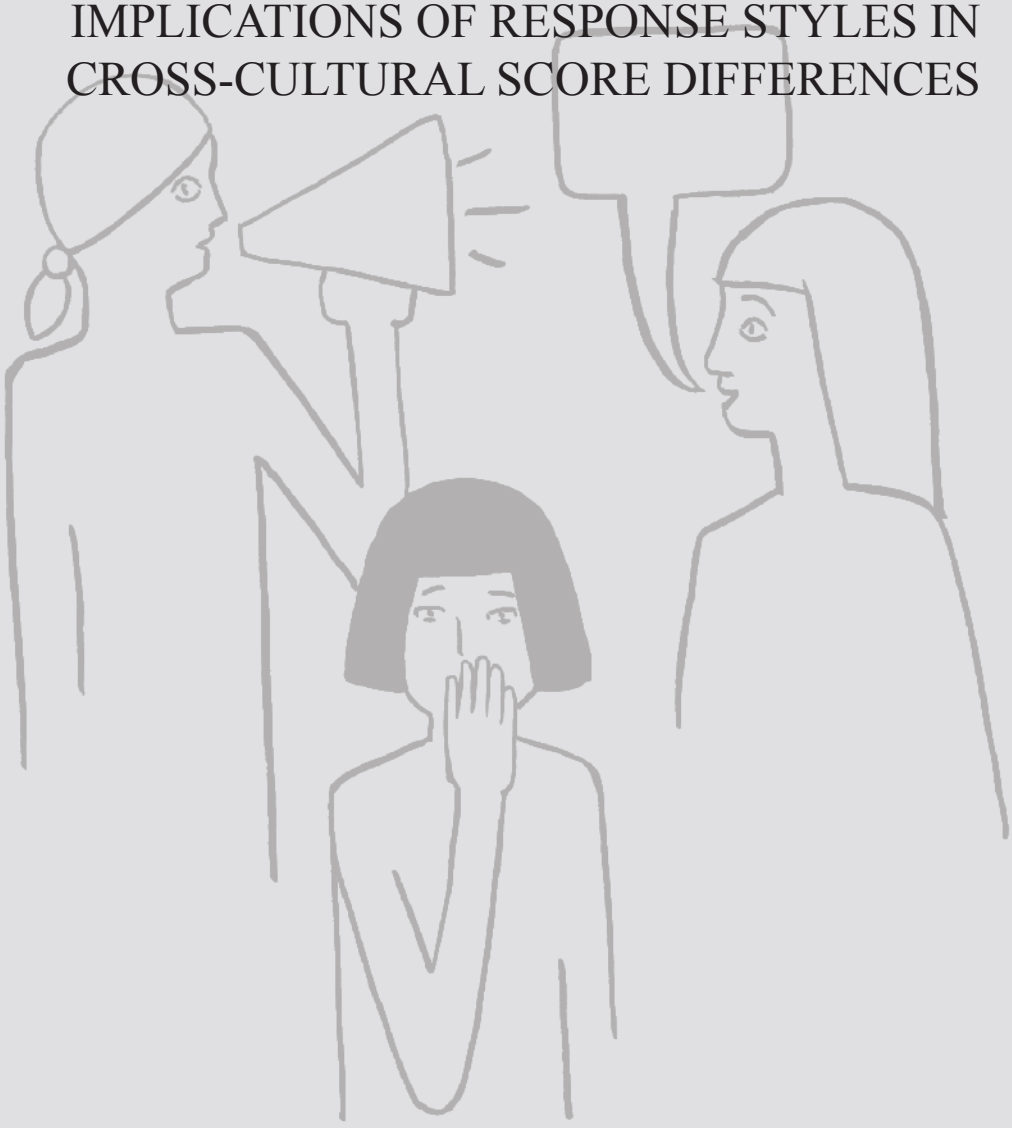
response formats in which different levels of uniform biases are embedded), and construct differences (e.g., the OPQ-based Big Five personality shared the working definitions of the five traits with the NEO). Mottus et al. (2012) found that country rankings of self-reported conscientiousness from Likert scales would change if response styles were taken into consideration, which demonstrated the strong uniform response biases resulting from the Likert scales. All in all, caution should be taken when interpreting the association of response styles or other constructs with the Big Five personality traits derived from different measures, especially when forced-choice format personality measures, such as the OPQ32r, can provide valuable additional information.

Limitations and Further Directions

The present study has a few limitations. Firstly, our response style indexes were constructed with items measuring other constructs, therefore it is difficult to ascertain that these response style indexes are adequately disentangled from the target constructs. However, the independent assessment and the high reliability in each response style alleviate these measurement concerns. Secondly, we operationalized ARS as the endorsement rate of the agreeing category with limited numbers of items, which may explain the very low loading of ARS on the GRS and its weak correlations with personality traits. However, irrespective of the various operationalizations of ARS, integrating different response styles into a general factor, which represents response amplification to moderation, helps to create consistency in findings. For these with a special interest in ARS, alternative data sources may be pursued. Items with dichotomized response options such as Agree/Disagree can be used to derive ARS indexes that are not confounded with ERS or MRS. Thirdly, the OPQ32r personality data were collected among working adults, which may pose some sample bias in generalizing the findings. Further efforts should improve the measurements of response styles, such as to develop direct measures of response styles or to locate external evidence to validate the indirect measures of response styles. When studying the underlying mechanism of response styles across cultures, it is also imperative to include unobtrusive measures such as HDI, measures with response formats that are resistant to response styles, and other techniques to effectively control for response styles (e.g., anchoring vignettes). In conclusion, the integration of specific response styles and the efforts to employ validity measures of different response formats are promising in deepening our understanding of the nature of response styles and ensuring valid cross-cultural comparisons.

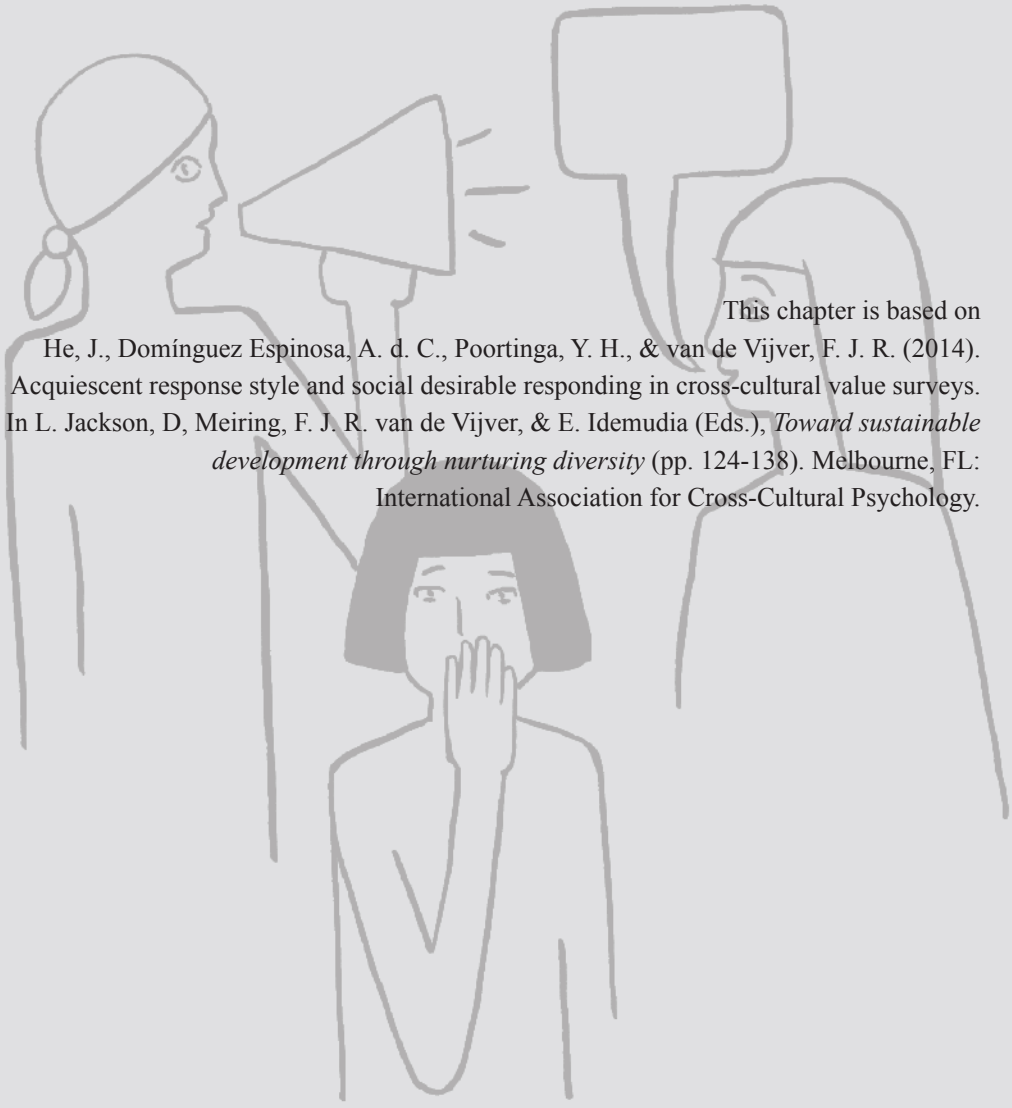
SECTION THREE

IMPLICATIONS OF RESPONSE STYLES IN CROSS-CULTURAL SCORE DIFFERENCES



Chapter 8

Acquiescent and Socially Desirable Response Styles in Cross-Cultural Value Surveys



This chapter is based on
He, J., Domínguez Espinosa, A. d. C., Poortinga, Y. H., & van de Vijver, F. J. R. (2014). Acquiescent response style and social desirable responding in cross-cultural value surveys. In L. Jackson, D, Meiring, F. J. R. van de Vijver, & E. Idemudia (Eds.), *Toward sustainable development through nurturing diversity* (pp. 124-138). Melbourne, FL: International Association for Cross-Cultural Psychology.

The present chapter reports two studies that are part of a larger research effort to systematically examine the effects of response styles on value scores from a cross-cultural perspective. In a first study we explored relationships among response styles, affluence, and values, and in the second study we mapped the effect of value score standardization on variance due to response styles. Value surveys can be vulnerable to various response styles. Response styles are found to be interrelated, and among the most frequently studied response styles are acquiescent response style (ARS: the tendency to agree to items irrespective of content) and socially desirable response style (SDR: the tendency to answer questions in a socially acceptable manner) (He & van de Vijver, 2013). In value studies, ARS and SDR seem to play an important role. Respondents' endorsement of a value item on a Likert scale may obscure the true standing on the value due to a tendency to react affirmatively rather than denial (Schwartz, 1992). Values are defined as desirable goals (Kluckhohn, 1951), and the desirability component embedded in instruments assessing values may create a tendency to react in a socially desirable manner (Fisher & Katz, 2000). In cross-cultural value surveys, issues become even more complicated, because differences in response styles and "true" value differences are confounded. Moreover, differences can arise not only at the individual level but also from cultural differences in values and response styles (e.g., T. P. Johnson, Kulesa, Cho, & Shavitt, 2005; P. B. Smith & Fischer, 2008).

Previous investigations have found that response styles at culture level are associated with a number of values; yet findings so far have made use of scattered data, and have reported in part conflicting findings (see below). In addition, a role for affluence in the relationship between response styles and values has been suggested, but seldom tested empirically. In Study 1, we address this issue with a large data set of ARS and SDR indexes at culture level. To account for the effects of response styles, especially ARS, within-subject or within-culture standardization of value scores has been proposed (Schwartz, 1992). Nonetheless, there is almost no empirical evidence to support the suggested removal of response style effects by standardization or to support the idea that this procedure offsets both ARS and SDR. We test the effects of score standardization of value scores from the Schwartz Value Survey in Study 2.

The contribution of our studies is that we use a large data set of ARS and SDR scores at culture level to examine their relationships with value scores and affluence, so as to validate previous findings with more stable estimates. In addition, we provide empirical evidence as to the effects of score standardization by comparing the correlation patterns of raw and standardized Schwartz value scores with ARS and SDR. In the next section, we first

review associations of ARS, SDR, affluence, and values that were reported previously. We then examine the findings of value score standardization on response styles and describe the research questions.

ARS, SDR, Affluence, and Values

ARS refers to the tendency to agree with propositions in general regardless of the content (Lentz, 1938). SDR is defined as the tendency of respondents to answer questions in a manner that will be viewed favorably by others (Crowne & Marlowe, 1960). It has been assumed that ARS and SDR are positively associated, because they exert joint impact on item responses (Ferrando & Anguiano-Carrasco, 2010), and both response styles are related to conformity (e.g., Gudjonsson & Young, 2011). Based on their shared affinity with conformity, we expect that ARS and SDR are positively related at culture level.

ARS at culture level (when operationalized as a weaker form of extreme response style) was found to be systematically associated with affluence-related indicators and cultural values in various multinational studies. Specifically, ARS showed negative associations with "hard" social indicators such as the Human Development Index (HDI), political freedom, and democracy (Meisenberg & Williams, 2008; van Dijk et al., 2009), and positive associations with "soft" indicators such as collectivism, hierarchy, and embeddedness (Harzing, 2006; P. B. Smith, 2004, 2011). There are also contradictory findings. For instance, Johnson, Kulesa, Cho, and Shavitt (2005) reported negative effects of power distance and uncertainty avoidance on ARS, whereas Harzing (2006) and P. B. Smith (2004, 2011) found a positive effect of power distance and a non-significant effect of uncertainty avoidance. The inconsistent results may be due to the fluctuation of ARS scores from different studies (e.g., number of countries covered, content of items from which these ARS indexes were constructed, and number of anchor points on the response scale). We replicate previous studies with ARS indexes compiled from more surveys involving various topics and answer scales, with a view to arrive at stable ARS measures.

There is limited empirical evidence on the culture-level correlates of SDR. The only culture-level study on the Lie scale of the Eysenck Personality Questionnaire, an SDR measure, showed substantial negative correlations with HDI, Gross National Product, political freedom, and democracy (van Hemert et al., 2002). It was also reported that the Lie scale was positively linked with cultural values of collectivism and power distance. Taking a similar approach to deriving culture-level scores of SDR, we expect the scores of the Marlowe-Crowne Social Desirability scale from different cultures to replicate these patterns.

To summarize, ARS and SDR are conceptually distinct yet related response styles that show similar patterns of correlations with affluence indicators. The correlates reviewed above imply that affluence may play a role in the relations between the two response styles and cultural values (e.g., P. B. Smith & Fischer, 2008). We investigate to what extent affluence serves as the common denominator in both response styles in Study 1.

Value Score Standardization

Score standardization refers to the procedure of transforming a set of raw scores to some other metric, usually by computing deviance scores (i.e., subtracting the mean of each score) or *z* scores (which corrects for individual differences in both means and standard deviations). The procedure has become a popular practice to adjust for response styles in data obtained from subjective evaluation of values, attitudes, and other psychological constructs using rating scales (Fischer, 2004). Score standardization takes different forms and these have different implications. Fischer (2004) distinguished within-subject, within-group, and within-culture standardization, and each may correct for differences in means, standard deviations, or both. It has been claimed that score standardization was effective in controlling for response styles (e.g., Hofstede, 1980; van de Vijver & Leung, 1997). At the same time, standardization may eliminate content-related differences or introduce new distortions (Dolnicar & Grün, 2007). There is no evidence that standardization removes all and only response style effects that jeopardize the validity of self-reported values.

In the Schwartz Value Survey, an explicit rationale and instruction for standardization are detailed for analysis of individual-level value types and for culture-level value types (Schwartz, 1992, 1999; Schwartz & Sagiv, 1995). For the former, within-subject standardization is recommended, which amounts to subtracting a respondent's mean score from each item score. The idea behind the standardization is that individuals can differ in their value preferences, but that one person cannot have more values than another. Therefore, differences in individual means across all value item ratings reflect scale use and not value substance. Value scores should be converted into scores that indicate the relative importance of each value in the participant's value system. On the basis of data from two countries Schwartz, Verkasalo, Antonovsky, and Sagiv (1997) suggested that within-subject standardization of value items mainly corrects for scale bias with only a small proportion of content-related variance being eliminated. Similarly, within-culture standardization is suggested for culture-level value types. Standardization here includes the correction of a

country's value type scores for the mean deviation of that country from the grand mean score of all countries (Schwartz, 2009b). Overall, it remains rather unclear whether and to which extent standardization affects response styles on value scores. Previous discussion has focused mainly on ARS and extreme responding (Dolnicar & Grün, 2007; Fischer, 2004), whereas SDR has been neglected. In Study 2 we study the correlation patterns of ASR and SDR with both raw and standardized value scores.

Summary of Research Questions

We seek to answer the following research questions:

1. What is the relationship at the culture level among ARS, SDR, and cultural values? (Study 1)
2. How much of the culture-level variance in the relationship of ARS and SDR with cultural values can be accounted for by affluence? (Study 1)
3. What are the effects when value scores are standardized to correct for ARS and SDR? (Study 2)

Study 1

To answer the first two research questions, we studied the associations of ARS, SDR, affluence, and cultural values. The study was carried out by means of secondary data analyses with large data sets on ARS and SDR at culture level.

Method

In this section, we describe the construction of multiple ARS measures, the meta-analytically derived SDR, the indicators for affluence, and the cultural value measures.

Culture-level ARS data. We located in the literature eight multinational surveys with Likert scales from which we could construct ARS indexes. Different response formats and scale characteristics were used in these surveys. Topics of the surveys range from social issues to learning motivations in different school subjects. More specifically, the ARS indexes were constructed from the European Values Survey (EVS, in 2008), Georgas et al.'s Family Value Study (in 2000), the European Social Survey (ESS, in 2008), the International Social Survey Program (ISSP; data sets in 2003, 2005, and 2006), the Program for International Student Assessment (PISA, in 2009), and the Trends in International Mathematics and Science Study (TIMSS, in 2007). Most datasets were published online; data of the Family

Value Study were provided by James Georgas. More information on the database is presented in Table 8.1.

Table 8. 1 ***Overview of Surveys for Constructing Acquiescence Indexes***

Survey	*N* of Items Utilized	Response Likert Points	*N* of Countries	*N* of Respondents
ESS (2008)	26	5	29	37,827
EVS (2008)	18	5	45	51,076
Family Value Study (2006)	36	7	27	5,194
ISSP 2003	6	5	29	37,768
ISSP 2005	8	5	16	22,653
ISSP 2006	7	5	33	48,641
PISA	4	4	14	83,361
TIMSS-Math	8	4	50	218,426
TIMSS-Science	8	4	29	140,419

Note. ESS = European Social Survey; EVS = European Values Survey; ISSP = International Social Survey Programme; PISA = Programme for International Student Assessment; TIMSS = Trends in International Mathematics and Science Study.

Computation of the ARS index. We derived ARS measures in two ways. The first, conventional operationalization was to select a variety of items that measure different constructs and use the same Likert answer scale. The proportion of answers expressing agreement is an index of ARS. For scales with bidirectional items (i.e., scales with both positively and negatively worded items), ARS indexes can be computed as the proportion of responses expressing agreement for positively worded items and for negatively worded items separately and the average can be taken as ARS scores (van Dijk et al., 2009).

In data obtained from ISSP (three waves), PISA, and TIMSS (two scales in TIMSS were used, one for motivation in learning math and the other one for science), we identified six scales with bidirectional items; hence, we calculated six ARS scores based on such items. For the other three surveys, we compiled three ARS indexes using the proportion of agreement responses (e.g., answers of 4 and 5 of the 5-point Likert scale). It should be noted that ARS indexes constructed this way (i.e., inclusion of the endorsement of the positive end of a scale) especially in unbalanced (positive vs negative) item sets can be viewed as a weaker form of extreme response style (He & van de Vijver, 2013).We first calculated the individual index for each respondent, and then averaged the index across members of the same country to obtain a country-level index. In total, we compiled nine ARS scores at country level, and the number of countries ranged from 14 to 50. We replaced the missing values of the ARS indexes across the surveys using EM imputation; the imputation was supported by the

nonsignificance of Little's MCAR test: $\chi^2(182) = 174.73$, $p = .64$ (Little & Rubin, 2002). We treated the nine ARS indexes as tau equivalent estimates (Lord & Novick, 1968), and used the mean as a global indicator of ARS. The value of Cronbach's alpha for this ARS scale was .87.

Culture-level SDR data. Culture-level SDR was derived from an instrument-based meta-analysis (van Hemert, 2003) using the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). This scale was chosen because of its popularity over the years and the prolific number of publications around the world. A total of 1052 publications from 1960 to 2011 which used the Marlowe-Crowne Social Desirability Scale or its various adapted forms were reviewed. Data from 153 journal articles, 4 unpublished theses, and 1 technical report were retained after excluding studies done with clinical, army, and inmate samples, and studies with "faking" or experimental instructions. Among all the studies, the scale was used in English (49.7%), French (10.1%), Spanish (8.2%), Chinese (6.9%), German (5.7%), Danish (3.8%), Portuguese (2.5%), Japanese, Swedish, Hebrew, Turkish and Arab/Iranian (1.9% each), Italian and Dutch (1.3% each), and Croatian, Greek, Norwegian, Indonesian, and Hindu (0.6% each).

Computation of the SDR index. Means and standard deviations of the Marlowe-Crowne Social Desirability Scale (33 items) were recorded for each study and aggregated to the culture level. For studies using different short or adapted forms of the scale, scores were converted using the following formula: $m \times [33 / (n \times r)]$, where m is the reported mean, n is the number of items, and r is the number of anchor options in the response scale. The means of the full original scale and those of the transformed scores were compared and no significant difference was found. Cronbach's alpha of the full scale was .76, and of the various short forms the reliability ranged from .76 to .82 (corrected with the Spearman-Brown formula). At the end, the mean scores of SDR at culture-level were obtained for 42 countries (for more details, see Domínguez Espinosa & van de Vijver, 2012).

Measures of affluence. The affluence level of a country was measured with the Human Development Index, literacy rates, and democratization.

The *Human Development Index* (HDI) is a composite measure of the average achievements of a country on three dimensions of human development: a long and healthy life (life expectancy), access to knowledge (enrolment in education), and standard of living (Gross Domestic Product per capita) (United Nations, 2010). Data were available for 194 countries.

Literacy rate of adults in a country (percentage of people aged 15 and above) were extracted from the World Data Bank for 130 countries (The World Bank, 2011) as an approximation of level of education.

Index of democratization, the entitlement of ideologically and socially different groups to compete for political power, was obtained from the Polyarchy Index of Democracy (Vanhanen, 2007). The average index from 1998 to 2000 was available for 184 countries.

Measures of cultural values. We compiled cultural value scores from five traditions.

(i) *Individualism, power distance, uncertainty avoidance, and masculinity* scores were taken from Hofstede (2009) ($N = 70$).

(ii) Schwartz cultural value dimension scores including *harmony, embeddedness, hierarchy, mastery, intellectual autonomy, affective autonomy,* and *egalitarianism* were taken from Schwartz (2009a). It should be noted that dimension scores here were standardized within-cultures ($N = 75$).

(iii) *Secularism* and *self-expression* scores were extracted from Inglehart's value dimensions (Inglehart et al., 2004) ($N = 80$).

(iv) The nine societal value dimensions including *power distance, uncertainty avoidance, institution collectivism, ingroup collectivism, gender egalitarianism, assertiveness, future orientation, performance orientation,* and *humane orientation* from the GLOBE leadership project were taken for 62 countries (House et al., 2004).

(v) Dimensions including *social cynicism, reward for application, social complexity, fate control*, and *religiosity* were taken from the Social Axioms scale for 39 countries (Leung & Bond, (2004).

Table 8. 2 ***Correlations of ARS and SDR with Affluence Factors***

	ARS	SDR
Human Development Index (HDI)	-.55*	-.66**
Literacy Rate	-.56*	-.55*
Democratization	-.51*	-.63**

Note. ARS = Acquiescent response style; SDR = Social desirability responding; HDI = Human Development Index.

Results

Correlation analyses were used to study the associations among the variables of interest. We describe the findings in two parts. First, we report the relationships of the two response styles and affluence indicators. Second, we compare the zero-order correlations and partial correlations (corrected for affluence) of the two response styles with cultural values.

ARS, SDR, and affluence. The correlation of the ARS and SDR was .48, $p < .01$, which indicated that the two response styles were rather strongly related at the culture level.

As shown in Table 8.2, both ARS and SDR showed strong negative correlations (-.51 to -.66, $p < .05$) with HDI, literacy rate, and democratization. Thus, affluence showed a consistent association with the two response styles, with less affluent countries tending to use more ARS and SDR.

Table 8. 3 ***Correlations of ARS and SDR with Cultural Value Scores from Five Research Traditions***

	ARS		SDR	
	Zero-order	HDI partialed out	Zero-order	HDI partialed out
Hofstede Values				
Power Distance	.14	-.20	.37*	-.15
Individualism	-.27*	.01	-.54**	-.09
Masculinity	.12	.10	.19	.20
Uncertainty Avoidance	-.07	-.14	-.04	-.08
Schwartz Values				
Harmony	-.35*	-.20	-.15	-.10
Embeddedness	.43*	.04	.56**	.13
Hierarchy	.36	.20	.30	.13
Mastery	.16	.18	-.15	-.03
Affective Autonomy	-.32*	-.05	-.50**	-.09
Intellectual Autonomy	-.49*	-.20	-.45*	-.07
Egalitarianism	.07	.31*	-.25	-.09
Inglehart Values				
Secularism	-.71**	-.59**	-.54**	-.06
Self-Expression	-.07	.36**	-.56**	.11
GLOBE-should be				
Power Distance	.12	.14	.02	.09
Uncertainty Avoidance	.36*	-.10	.32	-.10
Institution Collectivism	.42**	.29*	.25	-.06
Ingroup Collectivism	.31*	.28	-.13	-.08
Gender Egalitarianism	-.25	-.01	-.28	.04
Assertiveness	-.07	-.14	.19	.06
Future Orientation	.48**	.14*	.16	-.08
Performance Orientation	.33*	.26	.07	.03
Humane Orientation	-.15	-.23	-.20	-.18
Social Axioms				
Social Cynicism	.02	-.25	.48*	.29
Reward for Application	.61**	.27	.68**	.14
Social Complexity	-.36*	.11	-.60**	-.27
Fate Control	.31	-.28	.82**	.49*
Religiosity	.49**	.08	.71**	.32

Note. ARS = Acquiescent Response Style; SDR = Social Desirability Responding; HDI = Human Development Index. *$p < .05$. **$p < .01$.

The role of affluence in the correlations between response styles and values. Correlations of ARS and SDR with all the cultural value variables were computed, both with effects of HDI uncontrolled and controlled for. The two sets of correlations can be found in Table 8.3. As shown in the table, ARS was positively related to embeddedness, hierarchy, collectivism (GLOBE), uncertainty avoidance (GLOBE), future orientation, performance orientation, reward for application, and religiosity, and negatively related to individualism (Hofstede), harmony, secularism, and autonomy. However, nine out of the 13 significant correlations became nonsignificant after correcting for HDI. Leaving out the non-significant correlations in both zero-order and partial correlations, the average explained variance dropped from 17.02% to 6.24%.

We found positive correlations of SDR with power distance, embeddedness, social cynicism, reward for application, fate control, and religiosity. Negative correlations were found for individualism, autonomy, secularism, self-expression, and social complexity. Among the 12 significant correlations, only fate control stayed significant when HDI was controlled for. Leaving out the non-significant correlations in both zero-order and partial correlations, the average explained variance dropped from 31.47% to 4.64%.

Discussion

We set out to test the relationship between response styles and cultural values after statistically controlling for affluence. The findings support that affluence serves as a common denominator of relationships between the two response styles and cultural values. Correcting for affluence leads to a large reduction of the correlations between various psychological indicators and response styles. The reduction was larger for SDR than for ARS. Implications of the differential reductions in correlations are discussed in the General Discussion section.

Study 2

In this study, we explored the effects of value score standardization on ARS and SDR. We focused only on value scores from Schwartz Value Survey, for which there is a large data set so that we could construct value scores with raw and standardized data.

Method

Measures of ARS and SDR. The same measures of culture-level ARS and SDR as in Study 1 were used for the current analysis. Specifically, we had one global ARS index and one SDR index from the Marlowe-Crowne Social Desirability Scale.

Schwartz values of raw and standardized scores. With the original data of the Schwartz Value Survey collected by Schwartz and collaborators in 73 countries from 1988 to 2010, we were able to compute the seven culture-level value scores with both raw scores and within-culture standardized scores. Followed the coding manual (Schwartz, 2009b), the within-culture standardization was done in three steps: 1) calculate the mean of value ratings for all respondents from a country, 2) subtract the mean score from the scale mean (a score of 4 in this case), and 3) subtract the score difference obtained in Step 2 from each cultural value type score. In total, we had seven dimension scores in the raw and the standardized form.

Results

We correlated the response style measures with both raw and standardized scores of the Schwartz value types (Table 8.4).

Table 8. 4 ***Correlations of ARS and SDR with Raw and Standardized Schwartz Value Scores***

	ARS		SDR	
Culture-Level Type[2]	Raw Value Score	Standardized Value Score	Raw Value Score	Standardized Value Score
Harmony	.20*	-.36	.08	-.12
Embeddedness	.53*	.41*	.47**	.55**
Hierarchy	.51*	.32	.27	.20
Mastery	.58*	.21	.13	-.14
Affective Autonomy	.08	-.27*	-.43*	-.50**
Intellectual Autonomy	.04	-.46*	-.28	-.40*
Egalitarianism	.64*	.04	.06	-.18

Note. ARS = Acquiescent Response Style; SDR = Social Desirability Responding; HDI = Human Development Index. *$p < .05$. **$p < .01$.

[2] The entries for standardized value scores differ slightly from the corresponding entries in Table 3 because here a uniform set of 55 items in the SVS was taken for analysis.

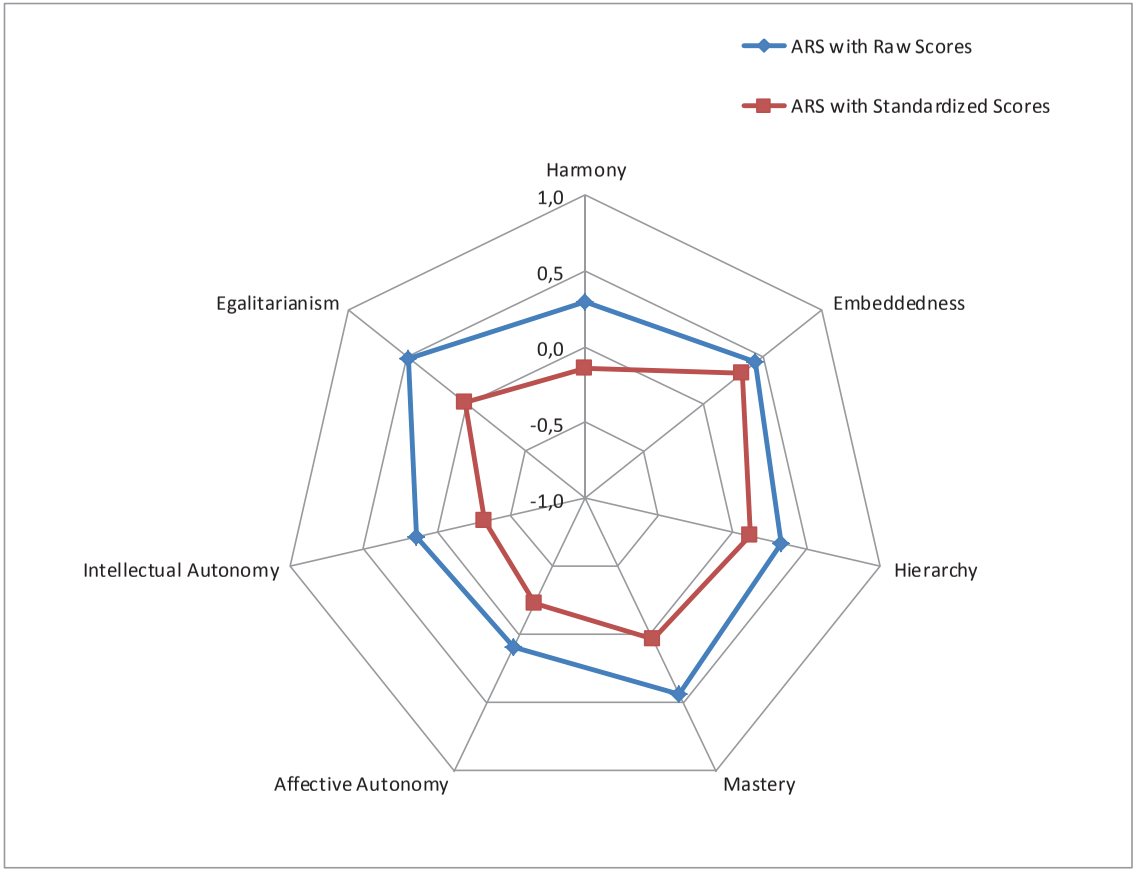


Figure 8. 1 Correlations of Acquiescence (ARS) with Raw and Standardized Culture-Level Schwartz Value Scores

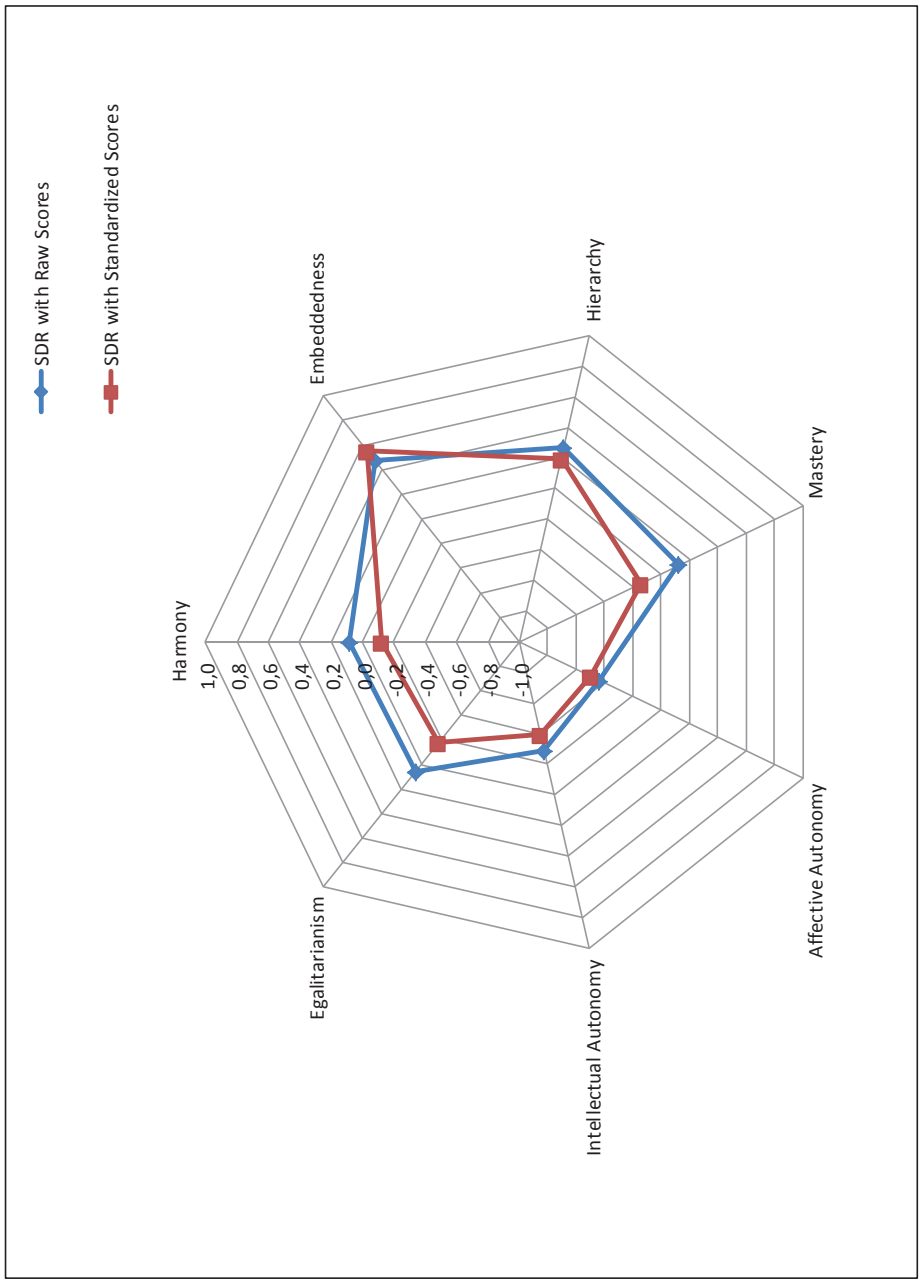


Figure 8. 2 Correlations of Social Desirable Responding (SDR) with Raw and Standardized Culture-Level Schwartz Value Scores

We found consistent, positive correlations of ARS with raw value scores. Because opposing values in the motivational continuum are conflicting and should show dissimilar correlations, the consistency of the positive correlations of all raw value scores with ARS suggests strong confounding of value scores and response styles. With standardized value scores, ARS was significantly associated to only a few value types, such as embeddedness versus autonomy, which pointed to the role of conformity. The variance explained in the value scores by ARS on average dropped from 18.90% to 10.52% as a consequence of score standardization. The shift of correlations between ARS and culture-level value types is shown in Figure 8.1.

The correlation patterns of SDR with raw and standardized value scores were very similar (Table 8.4). SDR showed positively associations with embeddedness (versus autonomy) in both sets of value scores. Standardization made the correlation slightly stronger. The explained variance increased from 10.04% to 13.94%. Still, the shift of patterning was rather limited and did not suggest a decrement in correlations after standardization as found for ARS. The shift of correlations between SDR and culture-level value types is presented in Figure 8.2.

Discussion

We compared the correlation patterns of ARS and SDR with raw and standardized value scores in the Schwartz value framework. We found that the correlation patterns of ARS with values were strongly affected by the standardization procedure, although not to the same extent for these value types, whereas the effects for SDR were more limited.

General Discussion

We address three main findings and some possible implications: (i) the correlation of ARS and SDR with affluence indicators; (ii) the pattern of correlations of these two response styles with value scores, with and without the effect of affluence partialed out; and (iii) shifts in the correlations between response styles and value scores as a consequence of value score standardization. We discuss the implications integrating results of both studies.

In Study 1 we found that the correlation between ARS and SDR at country level was rather strong and that both response styles showed substantial correlations with three affluence-related indicators. The correlation with affluence suggests substantial and systematic differences between countries. At face value the correlation between them would

seem to suggest that the two response styles address partly different aspects of the HDI variable. However, method effects may have played a role; the two styles were obtained through very distinct methods. ARS scores were derived from distributions of responses on Likert scales, whereas SDR scores were based on a separate measure (the Marlowe-Crowne scale). How the response style effects work out for value scores in five research traditions (Hofstede, Schwartz, GLOBE, Inglehart, and Social Axioms) is presented in Table 8.3. The table also shows that after correction for affluence the number of significant correlations decreases dramatically, especially for SDR, where only one of twelve entries remains significant. Perhaps more than any other of our findings this illustrates how response styles can influence findings of cross-cultural differences in the domain of values as they have been reported in numerous previous studies. In addition, the entries clearly suggest that there is differentiation in the size of the correlations. ARS and SDR probably should not be treated as having general effects; instead they affect differentially various values and value dimensions. This heterogeneity also holds across traditions; the GLOBE study which asked for values as they "should be" did not lead to any significant correlations with SDR across countries.

We found that ARS is mostly associated with values regarding "fitting in" or "normativeness" (secularism, institution collectivism, in-group collectivism). ARS is also positively correlated with embeddedness and negatively with autonomy, which equally points to "fitting in". This pattern of correlations is much weaker after correction for the effects of HDI.

As mentioned, with one exception significant correlations of SDR with value scores became insignificant after correcting for HDI. Low and insignificant correlations were found with various value scores known to be related to affluence (e.g., individualism and low power distance, secularism, and low autonomy), and for four of the five factors of the Social Axioms Scale, which measures generalized beliefs about people and events in the social world. Here we also found the only remaining significant effect. Ignoring this single exception, it can be concluded that affluence can account for most country differences in correlations between SDR and value scores.

Estimated effects of both ARS and SDR response styles are reduced strongly after correction for differences in HDI. However, standardization does not lead to large shifts in correlations for SDR, and the limited effects are difficult to interpret. For ARS standardization leads to substantial shifts in correlations, but these are more varied, at least for the value types distinguished by Schwartz (Schwartz, 1992, 1999) that we could analyze. In Table 8.4 some correlations for a given value type of standardized scores and raw scores even have an

opposite sign, while in other instances the correlation remains almost the same. Still, there is consistency in the sense that related values tend to show similar shifts. Figures 8.1 and 8.2 where value types are organized in a circle so that similar values are next to each other provide an illustration of this patterning for ARS.

In general terms there are three ways in which variance due to response styles can be interpreted. The first is to ignore such variance, effectively treating it as error variance. Although this option is chosen implicitly by authors when response style effects are not considered, this cannot be a viable option in cross-cultural research on values in view of the size of various correlations of response styles with other variables reported here. Ignoring the impact of response styles at country level would imply that affluence will be related to many values at country level, thereby obscuring real associations. The second way is to treat response style effects as systematic variance that is due to cross-cultural non-equivalence in measures used to assess value dimensions. Here response styles are seen as a validity threat and presumably the validity of scores will be improved if such non-target variance is corrected for. A study by Van Herk, Poortinga, and Verhallen (2004) can serve as an example. These authors found higher ARS scores for representative samples of respondents in countries from Southern Europe compared with samples from countries in Western Europe. At face value the higher scores expressed higher levels of product liking and purchase intentions. However, there was no evidence for corresponding national differences in volumes of actual purchases. In this case, response styles added non-target variance to scores, and should be viewed as cultural bias.

Standardization as recommended by Schwartz (2009b) amounts to a form of correction for general response tendencies that can be assumed to affect all scores equally. Apparently, standardization reduces effects of ARS on value scores. The better recovery of theoretically expected patterns (Table 8.4) and empirically support through the relationships with the other value measurements (Table 8.3) in the case of ARS seems to suggest that score standardization is an effective way to control for ARS in Schwartz Value Survey, at least partially. As mentioned, the entries in Table 8.4 suggest that ARS has differential effects for various value types. To counter the effects of SDR, standardization is a more remote option, since value score standardization does not affect the substantial correlations of SDR with the value type scores. If anything, the pattern of correlations is somewhat more sharply defined after standardization if SDR is seen as an expression of positive social values (e.g., embeddedness). Thus, a general correction to reduce or remove response style effects may be hard to justify.

The third way of dealing with cultural variance due to response styles is to approach this variance not as validity threat, but as providing information about the interpretation of cross-cultural differences in a target variable. For example, Smith (2004) interpreted differences in ARS as reflecting differences in collectivism. The correlations in Tables 8.3 and 8.4 suggest substantial relationships of SDR with individualism-collectivism, a dimension known to be highly correlated with affluence. Like individualism-collectivism SDR may be seen from this perspective as a psychological correlate of affluence. From this perspective any correction is likely to amount to overcorrection. The strong consequences of partialing out affluence-related variance underscore the need for a substantive interpretation; ARS and especially SDR are likely to contain information on the standing of individuals and cultures on value types and value dimensions. At the same time, it is conceptually an unsatisfactory state of affairs, if cross-cultural differences in response styles, in affluence, and in values appear as a mixture of interrelated variables. One way to disentangle the mixed effects is to view affluence as a moderator of the link between response styles and values. The affluence level of a country would then be a psychological indicator of the preferred national communication style (P. B. Smith, 2004, 2011), which provides a perspective in examining cross-cultural similarities and differences. Response styles are higher in less affluent countries, where people tend to endorse values pertaining to "fitting in".

It should be noted that there are some limitations in the present study. We constructed ARS indexes from different surveys with unbalanced numbers of countries available, which resulted in many missing values. Caution needs to be taken in generalizing the findings of correlations with ARS. In addition, we only studied the effects of value score standardization in one dataset (i.e., Schwartz Value Survey), while more value surveys with raw and standardized scores might be examined to test the generalizability of our findings. Last but not least, apart from the overlap in variance suggested by correlations in Table 8.3 and 8.4, there is also non-explained variance that requires further differentiation and interpretation.

Conclusions

In summary, in large data sets we found notable associations between ARS and SDR that cannot be ignored when cross-cultural score differences observed in various traditions of values research are interpreted. We examined effects of correction for affluence (HDI) and standardization. When a distinction is made between valid cross-cultural variance or target variance, and variance due to non-equivalence or bias of variables, evidence for both can be

pointed to. All in all, a complex picture has emerged from our two studies. In future research we intend to further tease apart variance components as mentioned in this chapter and extend the search to other types of response styles such as extremity and midpoint responding.

Chapter 9

Effects of a General Response Style on Cross-Cultural Comparisons: Evidence from the Teaching and Learning International Survey

This chapter is based on
He, J., & van de Vijver, F. J. R. (accepted). There are no effects of a General Response Style on cross-cultural comparisons: Evidence from the Teaching and Learning International Survey. *Public Opinion Quarterly*.

The Teaching and Learning International Survey (TALIS) is an OECD study that includes a principal survey and a teacher survey on various aspects of their working conditions and learning environments across countries (http://www.oecd.org/edu/school/talis.htm). The first round of the TALIS study was conducted in 2008 and the second round in 2013 (OECD, 2010, 2014). A recurrent issue in the analysis of the first round of the TALIS data involved the role of response styles, which refers to respondents' systematic tendency to answer questionnaire items on some basis other than the target construct (Paulhus, 1991). Response styles can influence the measurement structure, scale means, and relationships with other variables (van de Vijver & Leung, 1997; Welkenhuysen-Gybels et al., 2003). Analyzing data from the first round, Vieluf, Kunter, and van de Vijver (2013) found that one of the strongest correlates of country differences in teacher self-efficacy was acquiescent response style (ARS), the tendency to agree with items irrespective of the content. The study suggests that country mean differences in teacher self-efficacy cannot be taken at face value, as these are susceptible to response styles. We examine to what extent cross-cultural mean differences in various TALIS constructs, such as teacher self-efficacy, job satisfaction, and need for professional development, are influenced by response styles and to what extent cross-cultural differences in TALIS constructs are affected by corrections for response styles.

Besides ARS, three specific response styles have been frequently studied in surveys using self-reports: extreme response style (ERS; the tendency to frequently endorse the extremes in a response scale), midpoint response style (MRS; the tendency to frequently use the middle category or categories in a response scale), and socially desirable responding (SDR; the tendency to reply in a manner that is expected to be viewed favorably by others). ARS, ERS, and MRS are usually measured in an indirect manner (i.e., scores are derived from various items measuring other constructs), whereas SDR is directly measured using specific scales, such as the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). As explained below in more detail, one of the recurrent problems in the field of response styles (including SDR) is the lack of convergence of findings across studies (e.g., Greenleaf, 1992b; Helmes & Holden, 2003). Our study aims to integrate the conceptualization and measurement of response styles by using a generalized response style and to address its effects on cross-cultural mean differences in teacher self-reports in the second round of TALIS.

Measurement of Response Styles

Dimensionality of SDR. Different instruments of SDR typically do not yield converging factor structures (e.g., Helmes & Holden, 2003). The problem is exacerbated by the various conceptualizations of social desirability that have been proposed (Paulhus, 2002), such as the question of whether social desirability is primarily targeted at others (impression management) or the self (self-deception). However, even studies using the same scale do not always yield converging results. The most frequently used scale of SDR, the Marlowe-Crowne Social Desirability Scale, consists of 33 descriptions of highly desirable but rare and highly undesirable but common behaviors (Crowne & Marlowe, 1960). Initially conceptualized as unidimensional, this scale has been repeatedly found to be multidimensional. A two-factor solution has been most frequently reported. The two-factor solution distinguishes between *positive impression management* (the tendency to attribute socially desirable characteristics to oneself) and *negative impression management* (the tendency to deny socially undesirable characteristics). Examples of such studies can be found in the work by Millham (1974), Ramanaiah, Schill, and Leung (1977), and Loo and Loewen (2004).

Integration of specific response styles to a general factor. On the basis of their definitions, significant associations among ARS, ERS, MRS, and SDR can be expected. He and van de Vijver (2013) confirmed that a general response style factor (GRS) can integrate the four response styles, with SDR and ERS as positive indicators and ARS and MRS as negative indicators. At individual level, GRS is related to personality traits and values. At country level, GRS is negatively related to countries' socioeconomic development. It could be explained by the limited opportunities in life and the more restrictive norms in less developed countries which emphasize conformity and promote amplified self-expression (He, Bartram, et al., 2014). van de Gaer et al. (2012) suggested that response styles were associated with country-level educational achievement, with countries showing higher educational achievements using more moderating responses (e.g., East Asian).

Effects of Response Styles on Cross-Cultural Comparisons

There has been some controversy about the need to correct for response styles in survey research (D. H. Smith, 1967). On the one hand, it has been advocated that such response styles present distorted representations of participants' views, therefore correction is imperative. An example of this approach can be found in Eysenck and Eysenck's (1975) work, who proposed to interpret scores on personality scales only if a participant's score on a social desirability scale was below a pre-determined critical threshold. On the other hand, various

authors have argued that corrections for response styles do not have a sizeable impact on conclusions based on self-report scores. For instance, Ones, Viswesvaran, and Reiss (1996) reported that the validity of an instrument to predict job performance is not strongly influenced by a correction for response styles. Moreover, it has been proposed that response styles reflect preferences in communication styles (amplifying vs. moderating responses), which show differences at both individual and cultural level (He, Bartram, et al., 2014; P. B. Smith, 2004). Correcting for such preferences may remove valid individual and cross-cultural differences.

In cross-cultural contexts, the correction of response styles has shown mixed results. Diamantopoulos, Reynolds, and Simintiras (2006) claimed that response styles had an inconsistent impact on cross-cultural differences. In an eight-country study, Hoffmann, Mai, and Cristescu (2013) reported very small changes in correlations and mean differences after adjusting for response styles.

The Present Study

Our analysis of a large dataset of teachers' self-reports in the 2013 TALIS study adds to the literature in testing the factor structure of SDR cross-culturally, replicating the integration of response styles with a large-scale survey data, and empirically testing the extent to which country mean differences are influenced by response styles and whether statistical corrections for such styles would affect the nature and size of the country differences. Moreover, we extend the study by Vieluf et al. (2013) who analyzed correlations between acquiescence and teacher self-efficacy at individual and country level by addressing multiple response styles and estimating their influence on the size of the country differences in more TALIS constructs. We focus on three response styles: SDR, ERS, and MRS. We decided not to include ARS due to its incoherent operationalization in the literature. ARS is sometimes operationalized as a weak form of ERS by taking *agree* and *strongly agree* as its indicators, which limits the value of using both ERS and ARS. However, when ARS is operationalized as the proportion of *agree* responses (leaving out *strongly agree* responses), it is much more a measure of modesty as it then correlates negatively with ERS and positively with MRS (He & van de Vijver, 2013). We avoided this ambiguity by only using ERS and MRS.

Specially, we tested two hypotheses and addressed one research question:

Hypothesis 1: SDR has a positive and a negative impression management factor across countries.

Hypothesis 2: A general factor can be derived with SDR and ERS as positive indicators and MRS as a negative indicator.

In addition, we were interested in the influence of a statistical correction for response styles on observed cross-cultural differences. Given the lack of conclusive evidence as to the impact of such a correction, we did not test a specific hypothesis about its impact.

Method

Sample

The OECD contracted the International Association for the Evaluation of Educational Achievement Data Processing and Research Center to conduct the second round of TALIS survey in 2013. The target population of TALIS is lower secondary education teachers and leaders of mainstream schools. A stratified two-stage probability sampling design was applied in which teachers were randomly selected from the list of in-scope teachers for each of the randomly selected schools in each participating country. Each participating country selected its own time frame for survey administration, ranging from 12 days to 4 months, within the internationally prescribed time period of the end of the school year to ensure comparability. Both paper-and-pencil and online survey administration modes were used (OECD, 2014). The total participation rate among these lower secondary school teachers ranged from 51% in United States to 99% in Estonia.[3]

We used the 2013 TALIS data from national representative samples of lower secondary school teachers in countries that gave permission for inclusion of their data in the analyses. The SDR scale was an optional part of the TALIS survey that was not administered in each participating country. Data were available for 84% of the entire TALIS sample, amounting to a total of 94,260 teachers in 18 countries. Around 2% of the sample had incomplete responses on this scale. A missing value analysis was conducted with constructs of our interest (including items of this SDR scale, indicators of ERS and MRS, and scale scores of 17 core constructs), and missing completely at random was supported by Little's MCAR test: $\chi^2(21{,}069) = 19{,}033.58$, $p = 1.00$. We report analyses without respondents with missing data on the SDR scale; yet, we confirmed that analyses with imputed data yielded the same conclusions. In total, there were 76,887 teachers and the detailed information of their demographics is presented in Table 9.1.

[3] OECD, TALIS 2013 database (http://dx.doi.org/10.1787/888933048299)

Table 9. 1 ***Sample Characteristics***

Country	Sample Size	Mean Age	Proportion of Females	Experience - current school (years)	Experience - in total (years)
Abu Dhabi (UAE)	2187	38	.60	5	13
Brazil	12490	39	.67	7	14
Chile	1460	41	.62	10	15
Croatia	3478	42	.74	12	15
Estonia	2981	48	.83	14	21
Finland	5890	44	.70	10	15
France	2911	42	.66	9	16
Iceland	1176	45	.70	10	15
Korea	2783	43	.70	4	17
Latvia	4103	47	.88	15	22
Malaysia	3839	39	.72	7	14
Mexico	5015	42	.50	11	16
Poland	3694	42	.75	11	17
Portugal	6809	45	.73	11	20
Serbia	3549	43	.66	11	14
Slovak Rep.	3404	43	.81	12	18
Spain	9214	46	.59	9	18
United States	1904	42	.67	8	14
Total	*76887*	*43*	*.69*	*10*	*17*

The average age (in years) of teachers across countries ranged from 38 (Abu Dhabi, UAE) to 48 (Estonia). In most of the 18 countries, there were more female than male teachers. In all countries, teachers' highest level of educational attainment was on average ISCED Level 5A (i.e., Bachelor's degrees and Master's degrees from universities or equivalent institutions); 92% of the teachers obtained this level. Among all the teachers, years of experience in the current school ranged on average from 4 (Korea) to 15 (Latvia) years, and years of experience as a teacher in total ranged from an average 13 (Abu Dhabi, UAE) to 22 (Latvia) years. Age, percentage of females, and years of experience as a teacher at the current school and in total differed significantly across countries (Wilks' Lambda = .74, $F(85, 324{,}748) = 249.31$, $p < .01$, (partial) $\eta^2 = .06$).

Measures

SDR. SDR was measured with a short, adapted version of the Marlowe-Crowne Social Desirability scale. Ten items in the original scale were selected on two criteria: (1) items should not have an ambiguous meaning and (2) items should be appropriate in different cultural contexts. The shortened scale was expected to maintain its construct validity and

improve cross-cultural comparability. Given the focus of TALIS on the teacher workforce, item wording was adapted to a teaching and learning context. Five items were worded positively (e.g., "I help students and colleagues in trouble") and the other five items were worded negatively (e.g., "I have doubts about my ability to succeed as a teacher"). These items were translated into the languages of the respective countries; a rigorous translation verification process (which applied to all TALIS questionnaire items) was implemented to ensure consistency and comparability across the many contexts (OECD, 2014). The response options of this scale ranged from (1) *strongly disagree* to (7) *strongly agree*.

ERS and MRS. With the exception of the SDR scale, all Likert-scale items in the teacher survey had a 4-point response scale. We utilized these items to construct indexes of ERS and MRS. More specifically, 15 items from various constructs related to teacher feedback, personal beliefs in teaching, and school climate were randomly chosen to construct an ERS index. The average interitem correlation among these chosen items was .03, indicating sufficient heterogeneity in item content. The original responses were recoded as the presence of ERS (i.e., respondents' original responses on Likert scales of 1 and 4 as 1) or its absence (i.e., responses of 2 and 3 as 0). The reliability of the 15 recoded items was .73. The average ERS endorsement of the 15 items was then used as an indicator of ERS.

Another non-overlapping 15 items that used the 4-point Likert scale were randomly chosen (average interitem correlation = .02) to compute MRS scores. Existing scores were recoded as showing MRS (i.e., responses of 2 and 3 as 1) or not showing MRS (i.e., responses of 1 and 4 as 0). The reliability of the 15 recoded items was .70. The average endorsement was taken as the index of MRS. It should be noted that with the use of independent items to assess ERS and MRS, our indexes are based on independently derived figures that could show any correlation; yet, in practice we expect the indexes to be negatively correlated. This expectation was borne out; the individual-level correlation between ERS and MRS was -.78 ($p < .01$).

Table 9. 2 *Overview of the Core Constructs in TALIS*

Scale	Subscales(if any)	No of items	Sample Item	Cronbach's Alpha	Factor Scores from
Teacher efficacy	Efficacy in classroom management	4	Control disruptive behavior in the classroom	.82	Scalar
	Efficacy in instruction	4	Craft good questions for my students	.75	Scalar
	Efficacy in student engagement	4	Get students to believe they can do well in school work	.78	Scalar
Job satisfaction	Satisfaction with current work environment	4	I enjoy working at this school	.78	Metric
	Satisfaction with teaching profession	4	If I could decide again, I would still choose to work as a teacher	.81	Metric
Teacher cooperation	Exchange and coordination for teaching	4	Exchange teaching materials with colleagues	.72	Scalar
	Professional collaboration	4	Teach jointly as a team in the same class	.60	Scalar
Participation among stakeholders		5	There is a collaborative school culture which is characterized by mutual support	.85	Scalar
Teacher-student relations		4	In this school, teachers and students usually get on well with each other	.78	Scalar
Classroom disciplinary climate		4	Students in this class take care to create a pleasant learning atmosphere.	.85	Metric
Constructivist beliefs		4	My role as a teacher is to facilitate students' own inquiry	.71	Metric
Effectiveness in professional development		4	A group of colleagues from my school or subject group	.70	Scalar
Professional development in subject matter		5	Knowledge and understanding of my subject field(s)	.84	Metric
Professional development for diversity		6	Teaching in a multicultural or multilingual setting	.82	Metric

The Core Constructs in the Teacher Questionnaire. In the teacher questionnaire in TALIS, 17 core constructs (including 14 individual constructs and 3 composite constructs from some of these individual constructs) related to various aspects of the teaching profession were measured with 4-point Likert-scale items. An overview of these constructs is presented in Table 9.2. As documented in the technical report (OECD 2014), across all countries, all scales showed acceptable reliability (as shown in the values of Cronbach's α) and the standardized factor loadings of items for each scale were above .60. Measurement invariance of each scale across countries was tested in multigroup confirmatory factor analyses. Configural invariance (i.e., the same indicators loaded on the same latent variables across countries) was supported in all scales; metric invariance (i.e., factor loadings constrained to be equal across countries) was supported in most scales except a marginally acceptable fit on *Satisfaction with current work environment*, and scalar invariance (i.e., items constrained to have the same intercepts across countries) was rarely supported. However, in many cases, factor scores of the scale computed from scalar invariance showed very strong correlations with those based on metric invariance model, pointing to similar robustness of the cross-cultural differences between the two invariance models. For scales showing similarity in the two invariance models, factor scores of the scales were computed from the scalar invariance model; otherwise these were computed from the metric invariance model. The scale scores of each core construct across countries were produced by the weighted robust likelihood estimation. A linear transformation was applied to the scale scores, resulting in the distribution of scaled scores across all countries with a mean of 10 and a standard deviation of 2.

Results

We describe the results in three parts. We first report the measurement invariance test of the factors of the social desirability scale: positive and negative impression management. This test was carried out in a multigroup confirmatory factor analysis, which provided information on the level of comparability of scores across countries. We then report the extraction of GRS from the three specific response styles in an exploratory factor analysis, and its association with country-level socioeconomic and educational development to validate the GRS factor. Finally, we describe the impact of response style correction on country differences in the 17 core constructs. In this part, the correlations of GRS and the 17 core constructs at both individual level and country level were firstly computed to indicate whether

GRS had uniform or non-uniform associations with different constructs, followed by a multilevel analysis in which the country-level GRS predicted individual-level self-reports while controlling for individual differences in demographics; lastly, a series of Analyses of Covariance (ANCOVAs) were carried out to study the effect sizes and change of country rankings before and after GRS correction.

Measurement Invariance Analyses of the SDR Scale

Model testing. The two factors (positive and negative impression management) of the SDR measure were tested in a multigroup confirmatory factor analysis in AMOS 22 (Arbuckle 2006). We checked configural, metric, and scalar invariance. The model fit was evaluated by chi-square tests, Comparative Fit Index (CFI: acceptable above .90), the Tucker Lewis Index (TLI: acceptable above .90), and the Root Mean Square Error of Approximation (RMSEA: acceptable below .06) (Cheung & Rensvold 2002). Initial modification indices suggested correlated error terms between Item 8 "I help students and colleagues in trouble" and Item 9 "I admit when I do not know something if a student asks a question in class", and between Item 3 "I have doubts about my ability to succeed as a teacher" and Item 5 "I feel threatened by teachers who are very successful", possibly to account for the similarity in meaning of the two pairs of items. More specifically, item 8 and 9 refer to activities that are less central to teaching than are the other items. Item 3 and 5 refer to teacher activities whereas the other items of the negative impression management factor involve activities vis-à-vis pupils. Therefore, the error terms of the two pairs of items were correlated. Configural and metric invariance of this model were well supported. Scalar invariance was not supported (Table 9.3).

Table 9. 3 ***Model Fit of the Multigroup Confirmatory Factor Analyses***

Model	χ^2/df	CFI	TLI	RMSEA	$\Delta\chi^2/\Delta df$	ΔCFI	ΔRMSEA
Configural invariance	10.46**	.93	.91	.01			
Metric invariance	*10.97***	*.92*	*.90*	*.01*	*13.10***	*-.01*	*.00*
Scalar invariance	76.98**	.20	.26	.03	353.46**	-.72	-.02
Partial scalar invariance	39.31**	.63	.63	.02	183.37**	-.29	-.01

Notes. Most restrictive model with an adequate fit is printed in italics. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation. $^{**}p < .01$.

As scalar invariance of all items is often difficult to find in studies involving many countries, partial measurement invariance was examined. The constraints on the loadings and

intercepts of Item 9 “I admit when I do not know something if a student asks a question in class” and Item 10 “I am irritated by students who ask favors” were released because they showed the largest variations across countries. The fit of this partial scalar invariance model improved significantly, compared with the full scalar invariance models (Table 9.3). Standardized factor loadings of the metric invariance solution are presented in Figure 1. Loadings of all items were mostly significant and in the expected direction. The correlation between the two latent factors on average was -.50 (range: -.65, -.40).

Evaluating the effects of lack of invariance. Given the lack of support for scalar and partial scalar invariance, there is no statistical ground for comparing the observed score means. Invariance tests are sensitive to (among other things) sample size. In large-scale studies it is not always clear whether the lack of invariance is caused by major misspecifications of the model or minor misspecifications that are psychologically inconsequential and do not have severe consequences for the rank order of the country means (Byrne and van de Vijver 2010). To establish to what extent the size of the observed cross-cultural differences is affected by the lack of scalar invariance in the data, we examined the differences in the scalar invariance model and partial scalar invariance model using both a latent mean and an observed scale score approach (as detailed below).

First, with Brazil, which has the largest sample size among all countries, as the reference group, the *latent means* of both factors for each country were estimated using both the scalar and partial scalar invariance model. The correlation of the country scores of positive impression management from the two sets of estimations was 1, and that of negative impression management was .91, indicating very similar score patterns in the two models.

In the *observed mean* approach, the scale scores of positive and negative impression management were computed, each with five items or each with four items (as suggested in the partial invariance model). In the first case, the reliability (Cronbach’s alpha) was low, with values of .59 for the 5-item positive impression management scale, .48 for the 5-item negative impression management scale, and .61 if the two subscales were taken as one scale (with the items of the negative impression management scale reverse keyed). In the latter case (using four items), the scales showed a slightly better reliability, with values of .66, .52, and .65, respectively.

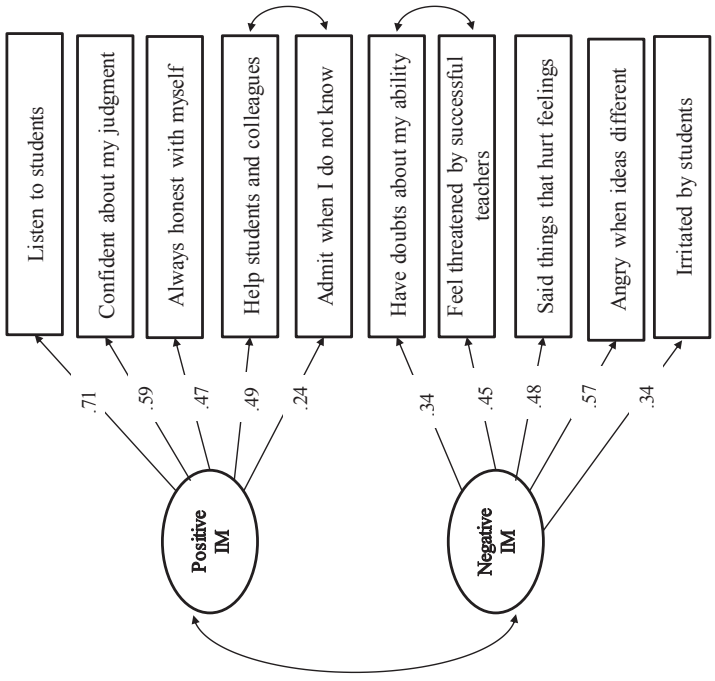


Figure 9. 1 *The Standardized Solution for the Social Desirability scale in the Metric Invariance Model*

The correlation between the scores of the 5-item and 4-item positive impression management at individual level was .91, and that of negative impression management was .90. At country level, the correlations were .80 and .75, respectively (all *p*s < .01). A Multivariate Analysis of Variance (MANOVA) was performed with the scale scores of positive and negative impression management as dependent variables, and country as the independent variable. The effect sizes were compared between the 5-item and 4-item solutions. The eta square values gauging the country differences dropped from .11 to .09, which suggested inconsequential differences in the size of cross-cultural differences.

To summarize, teacher's SDR data from these 18 countries demonstrated acceptable metric invariance and excellent partial metric invariance. Full scalar invariance was not supported by the fit statistics. However, the latent country means as well as observed scale scores based on scalar and partial scalar models did not show large differences and the size of cross-country differences were similar for the two models. Given the better fit of the partial invariance model, and the higher reliability of scales with only the invariant items, we decided to base the remaining analyses on the partial invariance model (i.e., using the 4-item positive and the 4-item negative impression management scale scores).

Integrating Specific Response Styles in a General Response Style Factor

A General Response Style factor (GRS), explaining 64% of the variance, was extracted in a Principal Component Analysis of the three response styles: SDR (the negative impression management items reversed and combined with the positive impression management scores), ERS and MRS. As expected, SDR (.51) and ERS (.91) loaded positively and MRS (-.91) loaded negatively on the factor. The presence of GRS suggests that the three response styles can be viewed as indicators of a common underlying tendency. GRS could be interpreted as a communication filter, which represents the tendency of either amplifying responses (e.g., using more SDR and ERS) or moderating responses (e.g., using more MRS).

To study its construct validity at country level, correlations were computed between GRS and the Human Development Index (HDI), a composite indicator of socioeconomic development of countries (United Nations, 2012), and country achievement in reading, mathematics, and science among 15-year-old students in the 2012 Programme of International Student Survey (OECD, 2013b). Negative associations were found for GRS with these country-level indicators (Table 9.4). So, countries with higher levels of socioeconomic development and higher levels of educational achievement are on average less inclined to show GRS than countries with lower levels of socioeconomic development and educational

achievement; these findings are in line with He, Bartram et al. (2014) and van de Gaer et al. (2012).

Table 9. 4 ***Correlation of the General Response Style (GRS) and Country-Level Indicators***

Country-Level Indicator (n = 18)	GRS
Human Development Index	-.23
PISA Reading Achievement Score	-.56*
PISA Mathematics Achievement Score	-.69**
PISA Science Achievement Score	-.63**

Note. *p < .05. **p < .01.

Implications of Corrections for GRS

To shed light on which constructs were more (or less) associated with response styles, we first investigated the correlations of GRS with core constructs in TALIS at both individual and country level (Table 9.5). The pan-cultural correlations (not correcting for country differences in mean scores) and pooled-within correlations (correcting for these differences) showed very similar patterning, thus only the pan-cultural correlations were reported here. Constructs related to teacher's efficacy including efficacy in classroom management, instruction, and student engagement showed the strongest correlations with GRS at both levels, followed by constructs related to different aspect of job satisfaction.

Table 9. 5 ***Correlations of the General Response Style (GRS) with the Core Constructs in TALIS at Both Individual and Country Level***

Scale	Individual Level	Country Level
Efficacy in classroom management	.33**	.49*
Efficacy in instruction	.41**	.73**
Efficacy in student engagement	.35**	.47*
Overall teacher efficacy	.39**	.61**
Satisfaction with current work environment	.41**	.77**
Satisfaction with teaching profession	.36**	.56*
Overall job satisfaction	.42**	.72**
Exchange and coordination for teaching	.13**	.15
Professional collaboration	.10**	.17
Overall teacher cooperation	.13**	.22
Participation among stakeholders	.15**	-.34
Teacher-student relations	.42**	.50*
Classroom disciplinary climate	.18**	-.01
Constructivist beliefs	.29**	.29
Effectiveness in professional development	.14**	.44
Professional development in subject matter	-.14**	-.33
Professional development for diversity	-.02**	.04

Note. *p < .05. **p < .01.

A set of multilevel analyses on the effects of country-level GRS on individual-level responses were carried out in HLM 6.06 (Raudenbush & Bryk, 2002) for two reasons. Firstly, the pancultural correlation analyses did not deal with data dependency across levels, thus it did not present a full picture of how country-level GRS (as validated in the previous section in its association with socioeconomic development and educational achievement) was associated with individual's self-reports. Secondly, the differences in sample demographics across countries were not controlled for in country-level correlational analysis. Each core construct (standardized across all countries) served as the dependent variable; age, gender, education, experience in the current school and in total served as individual-level predictors, and country-level GRS as the country-level predictor. As shown in Table 9.6, the variance of these core constructs that could be explained at country level ranged from 5% to 33%. Country-level GRS was positively associated with individual-level teacher's efficacy, job satisfaction, and teacher-student relationship.

The influence of GRS on the cross-cultural mean differences in these core constructs was assessed in a series of Analyses of Covariance (ANCOVAs). Each core construct served as a dependent variable, country as the independent variable, and GRS at individual level as a covariate. An important assumption of an ANCOVA is the homogeneity of regression coefficients of the covariate. If the regression coefficients (slopes) of the covariate are different across groups, the correction of the covariate could result in incorrect results. We tested the homogeneity of regression coefficients of GRS on each core construct across countries in multilevel analyses. The effect sizes of the interaction between individual- and country-level GRS on the standardized scores of the 17 core TALIS constructs were tested in a random slope and random intercept model. The standardized regression coefficients of the cross-level interaction ranged from -.05 to .01, which indicated very small effects; therefore, we concluded that the assumption of homogeneity of regression coefficients was not violated.

Table 9.7 presents the results of the ANCOVAs. The effect sizes before and after the correction were very similar for all constructs, and the correlations before and after the correction were higher than .90 in all cases. It seems that the response style correction did not affect the observed cross-cultural differences or the rank order of countries in a major way. However, we conclude with a caveat. Score corrections for these core constructs need to take into consideration if these constructs show scalar invariance themselves. We assumed that these constructs were not challenged by invariance issues. As the deviations in scalar invariance from metric invariance were minor in half of these scales (OECD, 2014), we expect that our analysis provides correct results.

Table 9. 6 ***Standardized Regression Weights in Multilevel Analysis***

Scale	ICC	Age	Gender (female)	Education	Experience - current school	Experience - in total	Country-Level GRS
Efficacy in classroom management	14%	-.04**	.03**	.00	.09**	.04**	.18*
Efficacy in instruction	23%	-.01	.06**	.01**	.06**	-.01	.34**
Efficacy in student engagement	26%	.01	.04**	.00	.05**	.00	.24**
Overall teacher efficacy	20%	-.01	.05**	.00	.07**	.01*	.27**
Satisfaction with current work environment	7%	.02**	.02**	-.03**	-.04**	.05**	.21**
Satisfaction with teaching profession	14%	.04**	.04**	-.02**	-.03**	.02**	.21*
Overall job satisfaction	10%	.04**	.03**	-.03**	-.04**	.04**	.23**
Exchange and coordination for teaching	22%	-.06**	.08**	.01**	.06**	.02**	.07
Professional collaboration	33%	-.06**	.03**	.01*	.06**	.03**	.09
Overall teacher cooperation	17%	-.07**	.06**	.01**	.07**	.02**	.09
Participation among stakeholders	5%	.04**	.00	-.03**	.02*	-.05**	-.08
Teacher-student relations	9%	.04**	.03**	-.02**	-.03**	.02**	.15*
Classroom disciplinary climate	8%	-.06**	.00	.00	.09**	.07**	.00
Constructivist beliefs	7%	-.01	.02**	.02**	.01	.00	.08
Effectiveness in professional development	8%	-.02	.05**	.02**	.05**	.00	.13
Professional development in subject matter	30%	-.04**	.00	-.01**	-.10**	.01	-.19
Professional development for diversity	23%	-.04**	.04**	.01	-.05**	.01	.02

Note. ICC = Intra-Class Coefficient; GRS = General Response Style. Gender was coded with 1 as female and 0 as male. *$p < .05$. **$p < .01$.

Table 9. 7 *Effects of Correcting for the General Response Style (GRS) in the Core TALIS Constructs*

Scale	Effect Size Before Correction	Effect Size After Correction	Correlation of Individual Corrected-Uncorrected Scores	Correlation of Country Corrected-Uncorrected Scores
Efficacy in classroom management	.11	.11	.95	.96
Efficacy in instruction	.21	.17	.91	.97
Efficacy in student engagement	.27	.26	.93	.98
Overall teacher efficacy	.18	.17	.92	.96
Satisfaction with current work environment	.06	.03	.91	.90
Satisfaction with teaching profession	.14	.13	.93	.95
Overall job satisfaction	.10	.07	.91	.92
Exchange and coordination for teaching	.24	.24	.99	1.00
Professional collaboration	.30	.27	1.00	1.00
Overall teacher cooperation	.13	.12	.99	.99
Participation among stakeholders	.04	.05	.99	.98
Teacher-student relations	.05	.06	.91	.88
Classroom disciplinary climate	.10	.11	.98	.98
Constructivist beliefs	.07	.06	.96	.93
Effectiveness in professional development	.05	.05	.99	.99
Professional development in subject matter	.21	.19	.99	1.00
Professional development for diversity	.23	.24	1.00	1.00

Discussion

The present study addressed the meaning and implications of response styles among teachers in 18 countries that participated in TALIS. These countries, ranging from Central and South America to East Asian countries, are known to differ considerably on ERS and MRS (P. B. Smith, 2011). Firstly, the analysis of invariance of the SDR scale showed strong support for the identity of the underlying factors, positive and negative impression management. Although scalar invariance was not fully supported, we found that comparisons of all items and items that are least affected by bias resulted in a similar pattern of cross-cultural differences, suggesting that despite the absence of a solid statistical basis for country-level comparisons of scores, such comparisons are likely to yield a fairly adequate picture.

Secondly, we replicated the findings of a GRS with SDR and ERS as positive indicators and MRS as a negative indicator. This GRS, representing the tendency of response amplification versus moderation has been confirmed using different instruments among different samples (e.g., He, Bartram, et al., 2014; He, van de Vijver, et al., 2014; He & van de Vijver, 2013). So, the GRS appears to be a stable construct, which holds promise to produce more stable results than studies of separate response styles.

Thirdly, we found that response styles do not affect TALIS core constructs in the same manner. The TALIS core constructs that were most related to response styles were teacher efficacy and teacher job satisfaction. There is evidence that response styles are triggered most by questions about personal domains, presumably because evaluation apprehension is then strongest (He & van de Vijver, 2014; van Dijk et al., 2009). Teachers may feel more personal involvement in evaluating their efficacy and job satisfaction than their professional development. The large number of constructs in TALIS allowed us to compare score corrections across domains that vary in GRS and personal involvement. The impact of score corrections was remarkably small across the board, both for constructs that showed very weak and very strong correlations with GRS. This result seems counterintuitive, even if it has been observed before (e.g., Hoffmann et al., 2013; Ones et al., 1996). The similarity of effect sizes of country differences before and after correction for response styles means that the correction leaves the country differences intact. The score rank order and the relative differences of country scores are not affected by these corrections. The same was observed at individual level. It seems fair to conclude that response styles, including SDR, ERS and MRS, have a detectable bearing on the TALIS data, notably at country level. However, "peeling off" the influence of response styles does not change the country order of means substantially.

Our study has some limitations. First, we only had 4-point scale items available to measure ERS and MRS, and we excluded ARS in the analysis. As it has been argued that response styles differ across response formats (Hui and Triandis 1989, 1985), further research should extract these response style indexes from new and more varied response formats. Secondly, scalar invariance was hardly confirmed in SDR and any of the core constructs used in the study. Yet, as argued before, it is not uncommon with large sample sizes in multiple countries. The effects of response styles on the measurement invariance of target constructs should be examined. Meanwhile, novel approaches to invariance such as Bayesian estimation (Muthen & Asparouhov, 2012) and sensitivity analysis of measurement differences (Oberski, 2014) may provide a more nuanced view on invariance in large-scale surveys. More experience with these procedures is needed (e.g., about specifying priors in the Bayesian estimation) to examine their suitability in large-scale cross-cultural surveys. Thirdly, we studied the effects of response styles on teaching related constructs, and further efforts should study various other constructs with sharper distinction on domains of personal relevance. To conclude, our study provides important empirical evidence of the general response style factor and its negligible correction effects in teachers' self-report in TALIS.

Chapter 10

The Motivation-Achievement Paradox in International Educational Achievement Tests: Towards A Better Understanding

This chapter is based on
He, J., & van de Vijver, F. J. R. (in press). The motivation-achievement paradox in international educational achievement tests: Towards a better understanding. In R. B. King & A. B. I. Bernardo (Eds.), *The psychology of Asian learners: A festschrift in honor of David Watkins*. Springer Asia.

In the Programme for International Student Assessment (PISA), a large-scale educational survey coordinated by the Organization for Economic Co-operation and Development (OECD), it is often found in all participating countries that students' self-report learning motivation positively predicts science, mathematics, and reading achievement (e.g., Chiu & Chow, 2010; Chiu & Zeng, 2008), which is in line with both Western and non-Western literature (Uguroglu & Walberg, 1979; Watkins, 2009). However, when scores are aggregated at country level and the correlation is computed between countries' average levels of motivation and achievement, a negative correlation is found (the terms culture and country are used interchangeably here). That is, East Asian countries, typically showing *high* scores on achievement in the PISA studies, such as China, Korea, and Japan, tend to have *low* scores on learning motivation. Such a paradox points to challenges in the comparability of data across countries. It deserves research attention to describe and explain the phenomenon. This chapter explores the methodological explanations and substantive interpretations of the paradox. We argue that differences in scale usage, embedded in Confucian cultural values (e.g., the emphasis on modesty in Asia), may fully or partially explain the negative association of motivation and achievement at country level.

In the following, we first review literature on intrinsic and extrinsic motivation in relation to academic achievement; we then describe a Chinese value pattern that we think is relevant for educational motivation and achievement. Next, we introduce three methods (i.e., response style correction, the overclaiming technique, and the anchoring vignettes technique) that have been proposed to account for scale usage differences across countries. We use PISA student data from 2012 to illustrate how these methods affect the association between motivation and achievement at country level. Finally, we discuss the substantive interpretation of the paradox.

Intrinsic and Extrinsic Motivation

People who are energized or activated toward an end are considered motivated. Two motivation orientations have been proposed (Ryan & Deci, 2000). *Intrinsic motivation* refers to doing an activity for its inherent satisfaction and enjoyment. It is often measured with self-report of interest in the activity *per se*. *Extrinsic motivation* refers to doing an activity in order to attain some separable outcomes, and is commonly measured with instrumental values of such an activity. It seems a universal finding that intrinsic motivation is positively related to high-quality learning and high achievement; however, studies on the association of extrinsic motivation and achievement showed mixed results, typically with a negative association in

Western cultures and a positive association in East Asian cultures (Wang & Guthrie, 2004; Watkins, 2000; Watkins, McInerney, Akande, & Lee, 2003). These differences are possibly due to cultural differences in learning for the sake of oneself or others (e.g., Bernardo & Ismail, 2010; R. B. King, McInerney, & Watkins, 2012b). Using the PISA student data from 2012, we explore whether both types of motivation are positively associated with academic achievement at individual level, and whether the associations at country level show a paradoxical pattern.

Chinese Values and Education

Where does the combination of high achievement and low motivation of Chinese learners come from? We argue that at least part of the explanation could come from Chinese values regarding effort, achievement, effort attribution, and views on the value of education. There is evidence that important people in social contexts of secondary school students (family members, school teachers, and peers) converge in their Confucianism-based view on the high *value of education* (e.g., Chen & Stevenson, 1995). However, other elements of Confucianism are also important. Compared to Western learners, Chinese learners are less inclined to have much confidence in their abilities and to *attribute* success to their own abilities (Mizokawa & Ryckman, 1990). Another important element is the focus on malleable components in performance, notably effort. The emphasis on the role of *effort, combined with the need for continuous self-improvement,* plays an important role. The Chinese saying "contentedness leads to loss, humility leads to gain" illustrates this drive for self-improvement through effort (KoonShingFrederick Leung, 2006). In addition, during socialization there is an emphasis on modesty and bragging is frowned upon (Xu, Zhang, & Hee, 2014). It is not surprising that self-concept scores of persons from Confucian countries tend to be low (Wilkins, 2004).

The description of this constellation of values, quite different from Western systems of values, is not meant to provide a causal mechanism for the combination of high scores on performance and low scores on motivation. After all, students who exert much effort do not always show a high achievement; furthermore, individual differences among Chinese students are expected to be considerable, thereby challenging the applicability of cultural values to each and every individual in that culture. However, the description clarifies how seemingly paradoxical associations can be interpreted from the backdrop of the Confucian values system. The description also clarifies that the often observed modesty bias and midpoint responding

by Chinese and other East Asian participants to surveys (Lee, Jones, Mineyama, & Zhang, 2002; Yu & Murphy, 1993) cannot be simply dismissed as erroneous responding.

Methods to Deal with Scale Usage Differences across Countries

The paradox in self-reported motivation and academic achievement in the PISA studies may be a result of differential scale usage by students in different countries. We review here three methods that can be applied in the PISA student data to account for the scale usage preferences. First, the systematic tendency of respondents to use certain portion of the response options independent of item content, known as response styles (Paulhus, 1991), are argued to be associated with cross-cultural variations in students' self-reports (e.g., Watkins, 1996; Watkins & Cheung, 1995). Chen, Lee, and Stevenson (1995) found that East Asian students who emphasize modesty tend to frequently use the middle categories in a Likert scale (i.e., midpoint response style), whereas Central and South American students are more inclined to use the extreme categories in such scales (i.e., extreme response style). Consequently, observed cross-cultural differences in self-report motivation scores may be partly or fully due to these response styles. Using Likert items measuring various constructs in the 2012 PISA study, we extract indexes of extreme and midpoint response styles for each student and their countries and investigate the role of culturally preferred response styles in the paradox.

A second indicator of scale usage differences is related to respondents' self-enhancement tendency. The strategy to maintain a positive image is sensitive to cultural contexts (Sedikides, Gaertner, & Vevea, 2005). It is argued that Westerners tend to self-enhance and show a better-than-average bias in the ability domain, whereas East Asians have a stronger self-criticism focus (Heine, Lehman, Markus, & Kitayama, 1999), which may explain the lower motivation scores in East Asian countries. The overclaiming technique is developed to capture the self-enhancement tendency independent of one's ability. It is expected that overclaiming is higher among Westerners compared with East Asians. Based on respondents' ratings of their knowledge of various persons, events, and products, some of which nonexistent, responses can be analyzed to indicate both self-enhancement bias (overclaiming knowledge of non-existent concepts) and accuracy (knowledge) (Paulhus, Harms, Bruce, & Lysy, 2003). This technique has been used in the PISA student questionnaire in 2012 (OECD, 2013a).

A third approach to deal with scale usage differences involves anchoring vignettes (G. King, Murray, Salomon, & Tandon, 2003). The anchoring vignettes technique refers to a

procedure in surveys that respondents rate themselves on certain traits, and they also rate several hypothetical persons described in written vignettes on the same traits. These hypothetical persons usually differ much in levels of the target traits. For example, one vignette describes a person with a very low level of conscientiousness, the second a medium level of conscientiousness, and the third a very high level of conscientiousness. The idea behind this technique is that by ensuring the presentation of fixed anchors with a predefined standing on the target construct, it becomes possible to understand how identical stimuli are linked to scale usage for each and every individual. Based on how the respondents rate themselves and how they rate these hypothetical persons, a rescaling of respondents' self-report to correct for interpersonal incomparability can be achieved (Hopkins & King, 2010). This technique has been applied in the PISA 2012 student questionnaire (OECD, 2013a).

To summarize, we set out to confirm the associations of self-report intrinsic and extrinsic motivation with math achievement within and between countries, and investigate whether response styles, overclaiming, and/or anchoring vignettes can explain this paradox.

Method

Procedure

The PISA student survey in 2012 assessed competencies of 15-year-olds in reading, mathematics, and science (with a focus on mathematics) in over 60 cultures. International experts from participating countries formed a committee, built the assessment frameworks, created and adapted items, and carried out extensive pre-tests to ensure the validity and reliability of measures (OECD, 2013a). Students were recruited through a stratified sampling procedure to represent the schools and the 15-year-old student populations of each country, and they took a background questionnaire and a subset of the cognitive test of different combinations that lasted two hours. The student questionnaire, data, manual, and the assessment frameworks are available on the OECD website (http://www.oecd.org/pisa/pisaproducts/).

Participants

In this study, 478,413 students between the ages of 15 years 3 months and 16 years 2 months from 64 countries were included (Table 10.1). There were 236,900 male students (49.5%) and 241,513 female students (50.5%).

Table 10. 1 *Sample and Country-Level Scores of Motivation, Response Styles, Overclaiming, and Anchored Motivation*

Country	Sample Size	Intrinsic Motivation	Extrinsic Motivation	ERS	MRS	Overclaim	Intrinsic Motivation (Anchored)	Extrinsic Motivation (Anchored)
Albania	4743	3.02	3.31	.40	.59	3.37	.47	.18
United Arab Emirates	11500	2.84	3.17	.41	.54	2.74	.30	.08
Argentina	5908	2.40	3.01	.37	.57	2.35	-.09	-.07
Australia	14481	2.36	3.08	.31	.64	2.26	.20	.26
Austria	4755	2.03	2.59	.40	.54	2.00	-.16	-.22
Belgium	8597	2.08	2.62	.30	.64	2.14	-.18	-.22
Bulgaria	5282	2.43	2.89	.34	.61	2.60	-.17	-.24
Brazil	19204	2.66	3.22	.34	.60	2.46	.13	.13
Canada	21544	2.33	3.11	.36	.58	2.41	.17	.26
Switzerland	11229	2.29	2.86	.35	.59	2.12	-.05	-.09
Chile	6856	2.51	3.14	.37	.54	2.46	.16	.15
Colombia	9073	2.67	3.18	.33	.60	2.59	.20	.17
Costa Rica	4602	2.52	3.12	.40	.51	2.33	.24	.19
Czech Rep.	5327	2.15	2.78	.28	.67	1.98	-.13	-.13
Germany	5001	2.19	2.80	.37	.56	2.02	.02	-.01
Denmark	7481	2.59	3.12	.30	.63	2.37	.28	.19
Spain	25313	2.15	2.87	.36	.60	1.74	-.11	-.06
Estonia	4779	2.27	2.93	.29	.68	2.07	.06	.05
Finland	8829	2.18	2.95	.31	.63	1.74	-.09	.05
France	4613	2.26	2.79	.33	.63	2.10	.03	-.10
United Kingdom	12659	2.38	3.16	.33	.62	2.08	.23	.30
Greece	5125	2.44	2.92	.36	.61	2.18	.02	-.08
Hong Kong-China	4670	2.51	2.74	.27	.67	2.04	.20	-.09
Croatia	5008	2.07	2.72	.32	.61	2.51	-.18	-.19
Hungary	4810	2.14	2.89	.31	.65	2.06	-.32	-.14
Indonesia	5622	2.92	3.21	.28	.70	2.90	.08	-.10
Ireland	5016	2.31	3.01	.33	.62	2.04	.26	.22
Iceland	3508	2.40	3.15	.38	.56	1.72	.20	.31
Israel	5055	2.41	3.11	.39	.54	2.14	.08	.14

Table 10. 2 *Sample and Country-Level Scores of Motivation, Response Styles, Overclaiming, and Anchored Motivation (Cont.)*

Country	Sample Size	Intrinsic Motivation	Extrinsic Motivation	ERS	MRS	Overclaim	Intrinsic Motivation (Anchored)	Extrinsic Motivation (Anchored)
Italy	31073	2.26	2.76	.34	.61	2.35	-.12	-.22
Jordan	7038	2.90	3.23	.47	.51	3.22	.11	-.10
Japan	6351	2.10	2.51	.31	.57	1.65	-.10	-.20
Kazakhstan	5808	2.96	3.21	.35	.64	3.02	.53	.24
Korea	5033	2.13	2.60	.27	.66	1.41	-.20	-.22
Liechtenstein	293	2.36	2.96	.39	.56	2.03	.10	.03
Lithuania	4618	2.36	3.09	.43	.54	2.16	-.02	.03
Luxembourg	5258	2.16	2.67	.38	.56	2.49	-.11	-.21
Latvia	4306	2.24	3.01	.27	.69	2.14	.03	.11
Macao-China	5335	2.39	2.72	.26	.69	2.19	.14	-.07
Mexico	33806	2.79	3.27	.35	.58	2.59	.34	.23
Montenegro	4744	2.30	2.70	.35	.60	2.87	-.20	-.34
Malaysia	5197	2.99	3.31	.32	.68	2.63	.31	.06
Netherlands	4460	2.02	2.64	.25	.71	2.20	-.17	-.12
Norway	4686	2.16	3.06	.33	.62	NA	-.19	.08
New Zealand	4291	2.36	3.12	.31	.64	2.23	.19	.31
Peru	6035	2.86	3.33	.29	.66	2.87	.49	.35
Poland	4607	2.15	2.81	.31	.66	2.64	-.26	-.23
Portugal	5722	2.36	3.10	.32	.62	2.44	.08	.16
Qatar	10966	2.75	3.12	.43	.54	2.94	-.03	-.18
Shanghai-China	5177	2.62	2.93	.30	.65	2.04	.61	.33
Romania	5074	2.66	2.46	.35	.64	2.88	-.21	-.73
Russian Federation	5231	2.50	2.84	.29	.67	2.56	.12	-.04
Singapore	5546	2.93	3.21	.34	.59	2.42	.68	.36
Serbia	4684	2.14	2.85	.34	.60	2.84	-.35	-.20
Slovak Republic	4678	2.12	2.65	.29	.67	2.10	-.41	-.44
Slovenia	5911	2.10	2.75	.30	.63	2.48	-.43	-.39
Sweden	4736	2.38	3.05	.32	.63	1.61	.12	.16

Table 10. 3 *Sample and Country-Level Scores of Motivation, Response Styles, Overclaiming, and Anchored Motivation (Cont.)*

Country	Sample Size	Intrinsic Motivation	Extrinsic Motivation	ERS	MRS	Overclaim	Intrinsic Motivation (Anchored)	Extrinsic Motivation (Anchored)
Chinese Taipei	6046	2.33	2.66	.31	.61	1.74	-.12	-.29
Thailand	6606	2.88	3.21	.28	.68	2.97	.26	.04
Tunisia	4407	2.75	3.20	.43	.54	2.44	.18	.08
Turkey	4848	2.62	2.94	.38	.58	2.43	.11	-.15
Uruguay	5315	2.48	3.06	.36	.59	2.08	.08	.07
USA	4978	2.35	3.02	.34	.60	2.38	.32	.30
Vietnam	4959	2.82	3.20	.21	.74	1.84	.39	.20
Total	478413	2.45	2.99	.34	.61	2.33	.08	.02

Measures

Students' scores on the two types of motivation and math achievement were derived from the data of student background questionnaire and the cognitive test, respectively. Measures of response styles, overclaiming, and anchoring vignettes were utilized with data collected with the student background questionnaire.

Intrinsic motivation to learn math was measured with a 4-item Likert scale on math interest. One sample item reads "I look forward to my mathematics lessons". *Extrinsic motivation* to learn math was measured with a 4-item scale on instrumental values of studying math (e.g., "Mathematics is an important subject for me because I need it for what I want to study later on"). The response options ranged from 1 (*strongly agree*) to 4 (*strongly disagree*). The reliability of the intrinsic motivation scale across the 64 countries ranged from .77 to .92, with a mean of .88, and that of the extrinsic motivation scale ranged from .79 to .92, with a mean of .88. Both scale scores were reverse-coded, so that a higher score on either scale indicated a higher motivation.

Students' *math achievement* was measured with different subsets of the cognitive test. A complex scheme of item administration, national calibrations, international scaling, and student score generation was applied to derive accurate population estimates of achievement scores (for details on the particular design and estimation procedure, consult the PISA technical report) (OECD, 2009, forthcoming). To put it simply, each student was administered only a subtest of the overall cognitive test to minimize test burden; however, by systematically varying items across student groups and using item response theory, it was still possible to derive information about the difficulty of all items. These cognitive data were then scaled using a Rasch model (which allows for missing data, due to the incomplete design), from which student math achievement scores were derived, denoted as plausible values. Plausible values are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured. Five plausible values of math achievement for each student were produced, thus standard analyses with math achievement should be performed on each of the plausible values and the results of each analysis should be combined (OECD, 2009; Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

Extreme response style was extracted from 15 randomly selected attitudinal items (excluding the two motivation scales) with 4-point response options in the student background questionnaire. Item content included subjective norms, self-efficacy, math anxiety, math self-concept, teacher support, perceived self-control, and so on. The average inter-item correlation

was .07, indicating sufficient item heterogeneity to capture the systematic response tendency rather than a substantive trait. The responses on these items were recoded with responses of 1 and 4 as 1, and other values as 0. The reliability of the recoded items ranged from .60 to .80 across countries with a mean of .71. The mean of the recoded items was taken as an index of extreme response style.

Midpoint response style was constructed in a similar manner as extreme response style. Another set of 15 nonoverlapping attitudinal items in the data were chosen (average inter-item correlation = .06) and recoded with responses of 2 and 3 in the 4-point scale as 1 and otherwise 0. The reliability of the recoded items ranged from .64 to .79, with a mean of .74. The mean of the recoded items was taken as an index of midpoint response style.

Three *overclaiming* items (i.e., items referring to concepts that do not exist) were administered along with items on the familiarity with math concepts in 63 countries (the items were not administered in Norway). The response option ranged from 1 (*never heard of it*) to 5 (*know it well, understand the concept*), and reliability ranged from .46 to .80, with a mean of .64. The mean rating of the three items was taken as an overclaiming score.

Two sets of *Anchoring Vignettes* (each with three vignettes) were applied to teacher support (on homework) and classroom management (on student interruption and teacher arrival time). The response options were from 1 (*strongly agree*) to 4 (*strongly disagree*). Comparing the relative standing of the self-report rating and the ratings of the vignettes, the anchoring vignettes were used to rescale the other constructs of the 4-point scales. Details of the procedure are described in a technical report (OECD, forthcoming). Here we use the anchored scale scores of intrinsic and extrinsic motivation provided in the student dataset. Scores of all the measures were first calculated for each student, and then averaged across students in the same country to derive country-level scores (Table 10.1).

Results

We describe the results in three parts. We first report the correlation of the two types of motivation and math achievement at both individual and country level to confirm the paradox. We then report the correlations of motivation and achievement at individual and country level taking into consideration of the effects of response styles, overclaiming, and the anchoring vignettes rescaling. Finally, we report a multilevel model to predict individual-level math achievement with individual-level motivation, and country-level response styles and overclaiming.

Association of Motivation and Achievement

The scale scores of intrinsic and extrinsic motivation were correlated with the five plausible values of math achievement at individual level in each country. The median correlations of intrinsic motivation and achievement (among the 5 correlations) ranged from -.13 to .42 across countries, with a median value of .17, and those of extrinsic motivation and achievement ranged from -.11 to .43, with a median value of .13, which indicated that at individual level, there is a trend of a positive association of both intrinsic and extrinsic motivation with achievement, although the effect sizes differed across countries (d'Ailly, 2003; Watkins, 2000).

The median correlations of intrinsic and extrinsic motivation with the five plausible values of achievement at country level were -.53 and -.52, respectively. The paradox in the association between motivation and achievement across levels was confirmed with data of math interest, instrumental values in math, and students' math performance in 64 countries.

Effects of Response Styles, Overclaiming, and Anchoring Vignettes on the Association of Motivation and Achievement

At individual level, the partial correlations of intrinsic motivation with math achievement controlling for extreme and midpoint response style across countries ranged from -.11 to .41, with a median correlation of .15, and those of extrinsic motivation with math achievement ranged from -.12 to .40, with a median correlation of .11. Controlling for overclaiming, the correlations of intrinsic motivation with math achievement across countries ranged from -.10 to .43, with a median value of .20, and those of extrinsic motivation ranged from -.12 to .46, with a median value of .15. Therefore, there were slight changes of correlations before and after controlling for response styles and overclaiming. Yet, the associations of both types of motivation with math achievement were still positive in either case. The correlations of rescaled intrinsic and extrinsic motivation based on anchoring vignettes with math achievement across countries ranged from .04 to .38 (median .26), and from .02 to .38 (median .20), respectively, indicating that after the rescaling, the positive associations between motivation and achievement largely remained at individual level (Table 10.2).

A visual inspection of country scores on response styles revealed that Asian countries such as Vietnam, Macao-China, Hong Kong-China, Korea, Indonesia, and Thailand ranked among the bottom ten on extreme response style, and the reversed ranking was found for

midpoint response style. Similarly, Asian countries including Korea, Japan, Chinese Taipei, and Vietnam ranked among the bottom ten on overclaiming, pointing to a strong response moderation tendency of these countries. A comparison of country rankings before and after anchoring vignettes rescaling also revealed the shift of Asian countries. For instance, Shanghai-China ranked 18th in the raw-score measures of intrinsic motivation, but it ranked 2nd after the rescaling.

Table 10. 4 ***Individual and Country-Level Correlation of Motivation and Achievement Controlling for Response Styles, Overclaiming, and Anchoring Vignettes***

Correlation with Achievement	Individual Level		Country Level	
	Intrinsic Motivation	Extrinsic Motivation	Intrinsic Motivation	Extrinsic Motivation
Zero-Order Correlation	.17	.13	-.53	-.52
Partialing Out Response Styles	.15	.11	-.45	-.44
Partialing Out Overclaiming	.20	.15	-.32	-.45
Using Anchored Scale Scores	.26	.20	-.11	-.04

Note. Individual-level correlations were median correlations across the five plausible values of achievement and then median correlations across countries; country-level correlations were median correlations across the five plausible values of achievement.

When the effects of country-level extreme and midpoint response styles, overclaiming and anchoring vignettes were taken into consideration, the median correlation of intrinsic motivation and achievement dropped from -.52 to -.45, -.32, and -.11, respectively, and that of extrinsic motivation and achievement dropped from -.53 to -.44, -.46, and -.04, respectively (Table 10.2).. In conclusion, these three methods, especially anchoring vignettes, could partially explain the negative association of motivation and achievement at country level.

Multilevel Analyses of Students' Achievement

To further explore the possible interactive effect of self-report motivation and culturally preferred scale usage on students' math achievement, a multilevel analysis was employed. We predicted students' achievement scores (i.e., the five plausible values of math performance) with students' motivation, the culturally preferred scale usage (i.e., extreme response style or overclaiming at country level), and the cross-level interaction. Midpoint response style was not used to avoid multicollinearity with extreme response style. All variables were standardized to *z* scores. Following Enders and Tofighi (2007), the individual-level predictor was centered around the group mean, and the country-level predictor around the grand mean. In each model, only one individual-level predictor, one country-level predictor, and their interaction were entered. The regression weights for all the five plausible

values of math achievement were essentially the same and the median values of the standardized regression weights are presented in Table 10.5.

Table 10. 5 ***Results of Multilevel Analyses***

Predictors	Median PV
Model 1	
Level 1: Intrinsic motivation	.15**
Level 2: Extreme Response Style	-.22**
Cross-Level Interaction	-.04**
Model 2	
Level 1: Extrinsic motivation	.13**
Level 2: Extreme Response Style	-.22**
Cross-Level Interaction	-.02**
Model 3	
Level 1: Intrinsic motivation	.14**
Level 2: Overclaiming	-.30**
Cross-Level Interaction	-.07**
Model 4	
Level 1: Extrinsic motivation	.13**
Level 2: Overclaiming	-.30**
Cross-Level Interaction	-.05**

Note. PV = Plausible values on achievement. $^{**}p < .01$.

As expected, both individual-level intrinsic and extrinsic motivation had a positive effect on students' math achievement. Country-level extreme response style and overclaiming had a moderate, negative association with students' math achievement. The cross-level interactions were significant, yet weak, indicating that culturally preferred scale usage may be a moderator in the associations of motivation and achievement. Specifically, in countries of high extreme response style and/or high overclaiming, the positive association of self-report motivation is less strong than in countries of low extreme response style and/or low overclaiming. For example, Mexico, a country with a high ranking on extreme response style, the individual-level correlation of intrinsic motivation and achievement was .06, whereas, in Korea, a country with a low ranking on extreme response style, the correlation was .42.

Discussion

We studied the paradoxical associations of two types of motivation (intrinsic and extrinsic) and math achievement among students from 64 countries. Correlational analyses showed that self-report motivation and achievement are positively associated at individual

level and negatively associated at country level. We found that controlling for individual scale usage differences through measures of response styles, overclaiming, and anchoring vignettes at individual level had differential impacts on the association of motivation and achievement. At country level, we first confirmed that Asian countries tend to have very low scores on extreme response style and overclaiming, and the rescaling based on anchoring vignettes changes the country rankings on the two types of motivation. We found that response styles, overclaiming, and anchoring vignettes can partially explain the negative association of motivation and achievement at country level. Besides, we found in a multilevel design that country-level extreme response style and overclaiming moderated the association of individual's self-report motivation and achievement. We focus our discussion on the interpretation of the paradox and further research directions.

In line with the correction effects of the three culturally preferred scale usage measures at country level, the Confucian values system emphasizing the value of education, modesty, self-criticism, efforts and continuous self-improvement can provide the conceptual backdrop of the seemingly paradoxical pattern of motivation and achievement. The importance of education, individual effort, and continuous self-improvement in Confucian cultures urges students to compete and perform well (e.g., R. B. King, McInerney, & Watkins, 2012a), whereas the cultural influence of modesty and self-criticism is imprinted on the scale use preferences as measured by response styles, overclaiming, and anchoring vignettes: Asian students use the extreme response style less and the midpoint response style more, tend to overclaim less, and use the response anchors in a way in line with the modesty bias (Chen et al., 1995). Overall, the paradoxical pattern at individual and country level can be in part interpreted by referring to the Confucian value system. Taking these values and culturally preferred scale usage embedded in these values into account, the paradoxical pattern of motivation and achievement is not that unexpected.

However, partialing out the effects of response styles, overclaiming, and anchoring vignettes did not completely reverse the country-level correlation, which points to the need to look for additional methodological explanations. Further research efforts should be made to advance our understanding of the paradox. For instance, intrinsic and extrinsic motivation in the PISA project were measured explicitly, which represent cognitively elaborated constructs, an implicit measure of motivation which represents a more primitive motivational system with affective experiences may be used and compared with the explicit measures (e.g., McClelland, Koestner, & Weinberger, 1989). The constructs of intrinsic and extrinsic motivation may be perceived differently across cultures, so that these constructs may not have

the same psychological meaning at individual and country level. Therefore, equivalence of the constructs across countries and across levels of analysis should be pursued (e.g., van de Vijver & Watkins, 2006). Furthermore, other sources of method and item bias may have played a role in the paradox, such as translation difficulties and administration differences across countries. At this stage, we urge researchers to be cautious in interpreting results of direct comparison of country scores of motivation from the aggregated self-reports.

Work on the Chinese learner and on Chinese education, with David Watkins' work as one of the driving forces behind it, has clearly expanded our insight in the link between culture and education. On the one hand, the work has shown that the picture of the Chinese learner as focusing on rote memory learning is a caricature; on the other hand, it has shown how values, deeply ingrained in a culture such as the Confucian value system, can have a major impact on various aspects related to education, ranging from motivation to achievement.

Chapter 11

General Discussion

It is remarkable that after six decades of research on response styles, the interpretation of these styles is still not clear-cut (e.g., Pauls & Stemmler, 2003; van Vaerenbergh & Thomas, 2013). This dissertation set out to contribute to the literature by addressing the systematic measurement, nomological network, and implications of response styles on survey responses from a cross-cultural perspective. This chapter begins with an overview of the main research questions and empirical findings, followed by a discussion of the theoretical and practical implications, and ends with recommendations for further research.

Research Question 1: Is there a General Response Style that can integrate specific response styles?

Empirical Findings. The integration of specific response styles was firstly tested at individual level in various ethnic groups in the Netherlands, using both indirect measures of acquiescence, extremity, and midpoint responding, and direct measures of acquiescence, extremity, midpoint responding, and social desirability (Chap 2 and Chap 3). A General Response Style was confirmed with positive loadings of extremity and social desirability, and negative loadings of acquiescence and midpoint responding. Then, this General Response Style was replicated using indirect measures of acquiescence, extremity, and midpoint responding with data from various waves of International Social Survey Program (Chap 6), and with meta-analytically derived scores of social desirability at culture level (Chap 7). It is interpreted as a communication filter that represents the tendency to moderate or amplify responses.

Implications. The confirmation of a General Response Style provides a framework to help create consistency in findings of response styles. In various studies using different instruments of specific response styles among different samples, this General Response Style has been replicated to indicate response amplification to moderation. Moreover, this general factor has been repeatedly identified with fewer than four specific response styles at both individual and cultural level, which points to its stability at both levels. Previous research has mainly focused on one or two specific response styles, and their operationalizations vary from study to study. Especially for social desirability, different scales tapping into different numbers of subdimensions are used (e.g., Paulhus, 2002), thus it is difficult to generalize the findings. Using the General Response Style factor that is extracted from nonoverlapping measures of specific response styles helps to integrate these specific styles in a parsimonious and systematic way, thereby probably inducing more consistency in findings. Meanwhile, each specific response style has some unique structure and meaning. For instance, a two-

dimension structure has been supported, distinguishing impression management and self-deception, or enhancement and denial, depending on the measures used. Therefore, these specific response styles should be investigated together with the General Response Style.

Research Question 2: What are the nomological network and cross-cultural variations of response styles?

Empirical Findings. At individual level, this General Response Style was found to be positively associated with age and negatively associated with education across datasets (Chap 2, Chap 6, and Chap 7). It was strongly related to positive personality traits and in particular the "Big One" personality (Chap 2). It was strongly related to positive life outcomes such as self-esteem and life satisfaction; moderately related to horizontal and vertical collectivism, horizontal individualism, and emotion appraisal; and weakly associated with political views (Chap 3). At cultural level, this General Response Style was negatively associated with socioeconomic development (Chap 6, Chap7, Chap 8 and Chap 9). It was positively related to values pertaining to "fitting in" and ambiguity avoidance (Chap 6), and country personality traits such as dominance, competitiveness, and data rationality (Chap 7). The nomological network of specific response styles is in line with their loadings on this General Response Style. However, it should be pointed out that each response style has some unique meaning beyond their shared meaning (Chap 3 and Chap 5).

At both levels, the domain specificity of response styles was confirmed: the higher the personal relevance, the higher the General Response Style (Chap 3, Chap 7, and Chap 9). For instance, the General Response Style at individual level had a stronger association with personally relevant constructs such as personality and cognition compared with less personally relevant constructs such as political views (Chap 3). This General Response Style at both levels showed a stronger association with teacher efficacy and job satisfaction compared with professional development (Chap 9).

Implications. There has been much debate on the nature of response styles. Are they nuisance, or are they substance (e.g., T. P. Johnson & van de Vijver, 2003; Schwartz et al., 1997)? This dissertation presented evidence leaning towards the substantive interpretation of response styles. In both individual and culture level, meaningful patterns of correlates have been found for the General Response Style and each specific response style. It indicates that response styles are embedded in individual dispositions and cultural characteristics. Most saliently, in the multilevel study of response styles and personality traits, even after the effects of country affluence and individualism have been controlled for, aggregated personality traits

that were measured with a forced-choice format instrument could still explain extra variance in response style, which provides strong evidence for the substance in response styles.

Given the evidence that response styles have some substantive meaning, it is important not to treat them as mere nuisance factors; instead these response styles represent valid individual and cultural differences. P. B. Smith (2009) proposed to make use of response styles to characterize national culture. Despite that response styles have been mostly treated as uninvited and unwanted variance, they have the potential to shed light on aspects of national culture (P. B. Smith, 2004, 2011). In this dissertation, the General Response Style is interpreted as a communication filter that is regulated by culture and that permeates self-reports.

Research Question 3: What are the implications of response style (correction) effects on cross-cultural score differences?

Empirical Findings. In general, response styles had a very limited effect on cross-cultural score differences. In comparisons among ethnic groups in the Netherlands, the correction of the General Response Style did not severely affect the effect sizes of cross-cultural differences in any construct (Chap 3). Correcting for self-presentation (which the General Response Style factor was one indicator) had weak effects on the intercorrelations among various psychological variables in longitudinal data of three waves (Chap 4). In cross-country comparisons of teachers' self-reports, the correction of the General Response Style hardly resulted in dramatic changes in country rankings, or the effect sizes of cross-cultural differences (Chap 9). Score standardization in value surveys did not seem to counter the effects of ARS and SDR to the same degree (Chap 8). The correction of response styles and overclaiming, and the rescaling of respondents' self-report scores based on anchoring vignettes could partially explain the negative association of motivation and academic achievement at country level, yet the effects of response style correction were milder than that of anchoring vignettes (Chap 10).

Implications. Findings on the meaningful patterns of response styles and the correction effects suggest that caution is needed in response style correction in cross-cultural comparisons. Especially in domains such as personality, positive life outcomes, cognition, values, and emotion regulation that show moderate to strong associations with response styles, correcting for response styles would get rid of valid individual and cultural differences (Fischer, 2004). Another consideration to refrain from response style correction is the small effects that response styles have on cross-cultural comparisons. There is no evidence that

response style corrections improve the validity of data, which indicates that response style correction is not imperative. In comparison of the correction of response styles and alternative measures to adjust for scale usage differences, anchoring vignettes seem to be able to better recover the comparability of aggregated data at culture level. More research is needed to link response style and anchoring vignettes.

Recommendations for Further Research

It is believed that progress in understanding the psychological meaning of response styles is contingent on attempts to unravel the mechanisms behind these styles in cross-cultural contexts. This dissertation has added to the literature regarding the systematic measurement, cross-cultural variations, and their correction effects on responses. Further research efforts can utilize multilevel analysis to study response styles with psychological variables such as personality and values concurrently at both individual and culture level (e.g., P. B. Smith & Fischer, 2008). To better understand the psychological processes underlying response style usage, mixed-method studies on response styles are encouraged (e.g., Morren et al., 2013). The investigations of correction effects of response styles in cross-cultural comparative studies can be extended to other existing large-scale international surveys such as the Program for International Student Assessment, the Program for the International Assessment of Adult Competencies, and European Social Survey, in order to generalize the findings.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27. doi: 10.1037/0033-2909.102.1.3

Arends-Tóth, J., & van de Vijver, F. J. R. (2009). Cultural differences in family, marital, and gender-role values among immigrants and majority members in the Netherlands. *International Journal of Psychology, 44*, 161-169. doi: 10.1080/00207590701545676

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences, 40*, 1235-1245. doi: 10.1016/j.paid.2005.10.018

Ayidiya, S. A., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly, 54*, 229-247. doi: 10.1086/269200

Bachman, J. G., & O'Malley, P. M. (1984). Black-White differences in self-esteem: Are they affected by response styles? *American Journal of Sociology, 90*, 624-639. doi: 10.2307/2779299

Bäckström, M. (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment, 23*, 63-70. doi: 10.1027/1015-5759.23.2.63

Barger, S. D. (2002). The Marlowe-Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment, 79*, 286-305. doi: http://dx.doi.org/10.1207/s15327752jpa7902_11

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *The European Journal of Public Health, 20*, 543-548. doi: 10.1093/eurpub/ckq052

Bartram, D. (2013a). A cross-validation of between country differences in personality using the OPQ32. *International Journal of Quantitative Research in Education, 1*, 182-211. doi: 10.1504/IJQRE.2013.056460

Bartram, D. (2013b). Scalar equivalence of OPQ32: Big five profiles of 31 countries. *Journal of Cross-Cultural Psychology, 44*, 61-83. doi: 10.1177/0022022111430258

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91*, 3-26. doi: 10.1037/0033-2909.91.1.3

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143-156. doi: 10.1509/jmkr.38.2.143.18840

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*, 186-204. doi: 10.1037/h0031474

Bernardo, A. I., & Ismail, R. (2010). Social perceptions of achieving students and achievement goals of students in Malaysia and the Philippines. *Social Psychology of Education, 13*, 385-407. doi: 10.1007/s11218-010-9118-y

Bye, H. H., Sandal, G. M., van de Vijver, F. J. R., Sam, D. L., Cakar, N. D., & Franke, G. H. (2011). Personal values and intended self-presentation during job interviews: A cross-cultural comparison. *Applied Psychology, 60*, 160-182. doi: 10.1111/j.1464-0597.2010.00432.x

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing, 10*, 107-132. doi: 10.1080/15305051003637306

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures*. (Doctor in Applied Economic Sciences Dissertation), Ghent University, Ghent.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131. doi: 10.1037/0022-3514.42.1.116

Chen, C., Lee, S.-y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*, 170-175. doi: 10.1111/j.1467-9280.1995.tb00327.x

Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development, 66*, 1215-1234. doi: 10.1111/1467-8624.ep9509180284

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi: 10.1207/s15328007sem0902_5

Cheung, M. W. L., & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling, 9*, 55-77. doi: 10.1207/s15328007sem0901_4

Chiu, M. M., & Chow, B. W. Y. (2010). Culture, motivation, and reading achievement: High school students in 41 countries. *Learning and Individual Differences, 20*, 579-592. doi: 10.1016/j.lindif.2010.03.007

Chiu, M. M., & Zeng, X. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction, 18*, 321-336. doi: 10.1016/j.learninstruc.2007.06.003

Clarke III, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review, 18*, 301 - 324. doi: 10.1108/02651330110396488

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Costello, A. B., & Osborne, J. W. (2005). Best practises in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*, 1-9.

Couch, A., & Keniston, K. (1960). Yea-sayers and nay-sayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology, 60*, 151-172. doi: 10.1037/h0040372

Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology, 33*, 401-415. doi: 10.1037/h0054677

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*, 3-31.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: studies in evaluative dependence*. New York, NY: John Wiley and Sons.

d'Ailly, H. (2003). Children's autonomy and perceived control in learning: A model of motivation and achievement in Taiwan. *Journal of Educational Psychology, 95*, 84-96. doi: 10.1037/0022-0663.95.1.84

Davis, R. E., Resnicow, K., & Couper, M. P. (2011). Survey response styles, acculturation, and culture among a sample of Mexican American adults. *Journal of Cross-Cultural Psychology, 42*, 1219-1236. doi: 10.1177/0022022110383317

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity, 44*, 761-775. doi: 10.1007/s11135-009-9225-z

de Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research, 45*, 104-115. doi: 10.1509/jmkr.45.1.104

Diamantopoulos, A., Raeynolds, N. L., & Simintiras, A. C. (2006). The impact of response styles on the stability of cross-national comparisons. *Journal of Business Research, 59*, 925-935. doi: 10.1016/j.jbusres.2006.03.001

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life scale. *Journal of Personality Assessment, 49*, 71-75. doi: 10.1207/s15327752jpa4901_13

Dolnicar, S., & Grün, B. (2007). Assessing analytical robustness in cross-cultural comparisons. *International Journal of Culture, Tourism and Hospitality Research, 1*, 140-160.

Domínguez Espinosa, A., Salas Menotti, I., & Reyes-Lagunes, I. (2008). Validez Concurrente de la Escala de Deseabilidad Social de Domínguez utilizando la Escala de Deseabilidad Social de Marlowe-Crowne [Concurrent validity of the Scale of Social Deseability of Domínguez using

the scale of social desirability of Marlowe-Crowne]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica, 25*, 125-139.

Domínguez Espinosa, A., & van de Vijver, F. J. R. (2012). *A cross-cultural meta-analysis of social desirable responding*. manuscript submitted for publication.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology, 17*, 245-260.

Dudley, N. M., McFarland, L. A., Goodman, S. A., Hunt, S. T., & Sydell, E. J. (2005). Racial differences in socially desirable responding in selection contexts: Magnitude and consequences. *Journal of Personality Assessment, 85*, 50-64. doi: 10.1207/s15327752jpa8501_05

Dwight, S. A., & Feigelson, M. E. (2000). A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educational and Psychological Measurement, 60*, 340-360. doi: 10.1177/00131640021970583

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *uropean Journal of Psychological Assessment, 16*, 20-30. doi: 10.1027//1015-5759.16.1.20

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138. doi: 10.1037/1082-989x.12.2.121

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.

Feather, N. T. (1991). Human values, global self-esteem, and belief in a just world. *Journal of Personality, 59*, 83-107. doi: 10.1111/j.1467-6494.1991.tb00769.x

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). Acquiescence and social desirability as item response determinants: An IRT-based study with the Marlowe–Crowne and the EPQ Lie scales. *Personality and Individual Differences, 48*, 596-600. doi: 10.1016/j.paid.2009.12.013

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35*, 263-282. doi: 10.1177/0022022104264122

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing, 17*, 105-120. doi: 10.1002/(sici)1520-6793(200002)17:2<105::aid-mar3>3.0.co;2-9

Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences, 40*, 873-884. doi: 10.1016/j.paid.2005.08.015

Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. *Cross-Cultural Research, 33*, 369-381. doi: 10.1177/106939719903300404

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96. doi: 10.1016/j.jrp.2005.08.007

Graziano, W. G., & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality, 70*, 695-728. doi: 10.1111/1467-6494.05021

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*, 176-188. doi: 10.2307/3172568

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56*, 328-351. doi: 10.1086/269326

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*, 348-362. doi: 10.1037/0022-3514.85.2.348

Gudjonsson, G. H., & Young, S. (2011). Personality and deception. Are suggestibility, compliance and acquiescence related to socially desirable responding? *Personality and Individual Differences, 50*, 192-195. doi: 10.1016/j.paid.2010.09.024

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences, 44*, 932-942. doi: 10.1016/j.paid.2007.10.034

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*, 243-266. doi: 10.1177/1470595806066332

He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology, 45*, 1028-1045. doi: 10.1177/0022022114534773

He, J., van de Vijver, F., J. R., Domínguez, A. d. C., & Mui, P. H. C. (2014). Acquiescent, extreme, and midpoint response styles: A multilevel study. *International Journal of Cross-Cultural Management*. doi: 10.1177/1470595814541424

He, J., & van de Vijver, F. J. R. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences, 55*, 794-800. doi: 10.1016/j.paid.2013.06.017

He, J., & van de Vijver, F. J. R. (2014). Integration and domain specificity of response styles: Towards a better understanding of a general response style. *manuscript submitted for publication*.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*, 766-794. doi: 10.1037/0033-295x.106.4.766

Helmes, E., & Holden, R. R. (2003). The construct of social desirability: one or two dimensions? *Personality and Individual Differences, 34*, 1015-1023. doi: 10.1016/S0191-8869(02)00086-7

Hoffmann, S., Mai, R., & Cristescu, A. (2013). Do culture-dependent response styles distort substantial relationships? *International Business Review, 22*, 814-827. doi: 10.1016/j.ibusrev.2013.01.008

Hofstede, G. (1980). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.

Hofstede, G. (2009). Dimension data matrix. Dimension Data Matrix Retrieved 03/02/2001 http://www.geerthofstede.eu/dimension-data-matrix

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*, 52-88. doi: 10.1177/1069397103259443

Hopkins, D. J., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly, 74*, 201-222. doi: 10.1093/poq/nfq011

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*, 296-309. doi: 10.1177/0022022189203004

Inglehart, R., Basafiez, M., Diez-Medrano, J., Halman, L., & Luijkx, R. (Eds.). (2004). *Human beliefs and values: A cross-cultural sourcebook based on the 1999-2002 values surveys*. Mexico City: Siglo Veintiuno Editores.

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence* Cambridge: Cambridge University Press.

Irwing, P. (2013). The general factor of personality: Substance or artefact? *Personality and Individual Differences, 55*, 234-242. doi: http://dx.doi.org/10.1016/j.paid.2013.03.002

Jarvis, W. B. G., & Petty, R. E. (1996). The need to evaluate. *Journal of Personality and Social Psychology, 70*, 172-194. doi: 10.1037/0022-3514.70.1.172

Johnson, J. A. (1981). The 'self-disclosure' and 'self-presentation' views of item response dynamics and personality scale validity. *Journal of Personality and Social Psychology, 40*, 761-769. doi: 10.1037/0022-3514.40.4.761

Johnson, T. P., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response syles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*, 264-277. doi: 10.1177/0022022104272905

Johnson, T. P., & van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195-204). Hoboken, New Jersey: John Wiley & Sons.

Just, C. (2011). A review of literature on the general factor of personality. *Personality and Individual Differences, 50*, 765-771. doi: 10.1016/j.paid.2011.01.008

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2003). Enhancing the validity and cross-cultural comparability of measurement in survey research. *The American Political Science Review, 97*, 567-583. doi: 10.2307/3593024

King, R. B., McInerney, D. M., & Watkins, D. A. (2012a). Competitiveness is not that bad…at least in the East: Testing the hierarchical model of achievement motivation in the Asian setting. *International Journal of Intercultural Relations, 36*, 446-457. doi: 10.1016/j.ijintrel.2011.10.003

King, R. B., McInerney, D. M., & Watkins, D. A. (2012b). Studying for the sake of others: the role of social goals on academic engagement. *Educational Psychology, 32*, 749-776. doi: 10.1080/01443410.2012.730479

Kluckhohn, C. (1951). Values and value-orientations in the theory of action: An exploration in definition and classification. In T. Parsons & E. Shils (Eds.), *Toward a general thoery of action* (pp. 388-433). Cambridge, MA: Harvard University Press.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. doi: 10.1002/acp.2350050305

Lalwani, A. K., Shavitt, S., & Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology, 90*, 165-178. doi: 10.1037/0022-3514.90.1.165

Lalwani, A. K., Shrum, L. J., & Chiu, C.-y. (2009). Motivated response styles: The role of cultural values, regulatory focus, and self-consciousness in socially desirable responding. *Journal of Personality and Social Psychology, 96*, 870-882. doi: 10.1037/a0014622

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a likert scale. *Research in Nursing & Health, 25*, 295-306. doi: 10.1002/nur.10041

Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement, 65*, 140-154. doi: 10.1177/0013164404267285

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin, 35*, 659.

Leung, K. (2006). Mathematics education in East Asia and the West: Does culture matter? In F. S. Leung, K.-D. Graf & F. Lopez-Real (Eds.), *Mathematics Education in Different Cultural Traditions-A Comparative Study of East Asia and the West* (Vol. 9, pp. 21-46): Springer US.

Leung, K., & Bond, M. H. (2004). Social Axioms: A model for social beliefs in multicultural perspective. *Advances in Experimental Social Psychology, 36*, 119-197. doi: 10.1016/s0065-2601(04)36003-x

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*, 131-141. doi: 10.1111/j.1468-2389.2006.00339.x

Li, A., & Reb, J. (2009). A cross-nations, cross-cultures, and cross-conditions analysis on the equivalence of the Balanced Inventory of Desirable Responding. *Journal of Cross-Cultural Psychology, 40*, 214-233. doi: 10.1177/0022022108328819

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology, 83*, 854-864. doi: 10.1037/0022-3514.83.4.854

Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe–Crowne Social Desirability Scale. *Journal of Applied Social Psychology, 34*, 2343-2352. doi: 10.1111/j.1559-1816.2004.tb01980.x

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores* Reading, Massachusettes: Addison-Wesley.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology, 23*, 498-509. doi: 10.1177/0022022192234006

Marshall, R., & Lee, C. (1998). A cross-cultural, between-gender study of extreme response style In B. G. Englis & A. Olofsson (Eds.), *European Advances in Consumer Research* (Vol. 3, pp. 90-95). Provo, UT: Association for Consumer Research.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690-702. doi: 10.1037/0033-295X.96.4.690

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 105-125). New York, NY: Kluwer Academic Publisher.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888. doi: http://dx.doi.org/10.1037/0022-006x.51.6.882

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures, P. (2005a). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407-425. doi: 10.1037/0022-3514.89.3.407

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures, P. (2005b). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547-561. doi: 10.1037/0022-3514.88.3.547

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*, 1539-1550. doi: 10.1016/j.paid.2008.01.010

Meiser, T., & Machunsky, M. (2008). The personal structure of personal need for structure: A mixture-distribution Rasch analysis. *European Journal of Psychological Assessment, 24*, 27-34. doi: 10.1027/1015-5759.24.1.27

Messick, S. (1991). Psychology and methodology of response styles. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach.* (pp. 161-200). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messick, S., & Jackson, D. N. (1961). Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin, 58*, 299-304. doi: 10.1037/h0043979

Millham, J. (1974). Two components of need for approval score and their relationship to cheating following success and failure. *Journal of Research in Personality, 8*, 378-392. doi: 10.1016/0092-6566(74)90028-2

Minkov, M. (2007). *What makes us different and similar: A new interpretation of the World Values Survey and other cross-cultural data*. Sofia, Bulgaria: Klasika i Stil.

Minkov, M. (2009). Nations with more dialectical selves exhibit lower polarization in life quality judgments and social opinions. *Cross-Cultural Research, 43*, 230-250. doi: 10.1177/1069397109334956

Minkov, M., & Blagoev, V. (2009). Cultural values predict subsequent economic growth. *International Journal of Cross Cultural Management, 9*, 5-24. doi: 10.1177/1470595808101153

Mizokawa, D. T., & Ryckman, D. B. (1990). Attributions of academic success and failure: A comparison of six Asian-American ethnic groups. *Journal of Cross-Cultural Psychology, 21*, 434-451. doi: 10.1177/0022022190214003

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2012). Response strategies and response styles in cross-cultural surveys. *Cross-Cultural Research, 46*, 255-279. doi: 10.1177/1069397112440939

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (2013). Exploring the response process of culturally differing survey respondents with a response style: A sequential mixed methods study. *Field Methods, 25*, 162-181. doi: 10.1177/1525822x12453535

Mõttus, R., Allik, J., Realo, A., Rossier, J., Zecca, G., Ah-Kion, J., . . . Johnson, W. (2012). The effect of response style on self-reported conscientiousness across 20 countries. *Personality and Social Psychology Bulletin*. doi: 10.1177/0146167212451275

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality, 41*, 1213-1233. doi: 10.1016/j.jrp.2007.02.003

Muthen, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335. doi: 10.1037/a0026802

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality, 77*, 261-286. doi: 10.1111/j.1467-6494.2008.00545.x

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*, 263-280. doi: 10.1002/ejsp.2420150303

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass, 2*, 842-860. doi: 10.1111/j.1751-9004.2007.00059.x

Oberski, D. L. (2014). Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models. *Political Analysis, 22*, 45-60. doi: 10.1093/pan/mpt014

OECD. (2009). Analyses with plausible values *PISA data analysis manual: SPSS* (2nd ed.). Paris: OECD publishing.

OECD. (2010). *TALIS 2008 technical report*. Paris: OCED Publishing.

OECD. (2013a). *PISA 2012 Assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.

OECD. (2013b). *PISA 2012 results in focus*. Paris: OECD Publishing.

OECD. (2014). *TALIS 2013 technical report*. Paris: OECD publishing.

OECD. (forthcoming). *PISA 2012 technical report*. Paris: OECD Publishing.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679. doi: 10.1037/0021-9010.81.6.660

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609. doi: 10.1037/0022-3514.46.3.598

Paulhus, D. L. (1991). Measurement and control of response biases. In J. Robinson, P. Shaver & L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 17-59). San Diego, CA: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaune.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890-904. doi: 10.1037/0022-3514.84.4.890

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307-317. doi: 10.1037/0022-3514.60.2.307

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*, 263-275. doi: 10.1016/s0191-8869(02)00187-3

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*, 531-544. doi: 10.1177/014920638601200408

Preisendörfer, P., & Wolter, F. (2014). Who is telling the truth?: A validation study on determinants of response behavior in surveys. *Public Opinion Quarterly, 78*, 126-146.

Press, J. E., & Townsley, E. (1998). Wives' and husbands' housework reporting: Gender, class, and social desirability. *Gender and Society, 12*, 188-218. doi: 10.2307/190530

Ramanaiah, N. V., Schill, T., & Leung, L. S. (1977). A test of the hypothesis about the two-dimensional nature of the Marlowe-Crowne Social Desirability scale. *Journal of Research in Personality, 11*, 251-259. doi: 10.1016/0092-6566(77)90022-8

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*, 53-61. doi: 10.1016/j.jrp.2009.10.005

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2 ed.). Newbury Park, CA: Sage.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *Journal of clinical psychology, 38*, 119-125.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*, 617-640. doi: 10.1111/1467-6494.694157

Rokeach, M. (1973). *The nature of human values*. New York, NY: Free Press.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Prinsceton, NY: Princeton University Press.

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior, 25*, 189-197.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*, 142-151. doi: 10.3102/0013189x10363170

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67. doi: 10.1006/ceps.1999.1020

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177. doi: 10.1037/1082-989X.7.2.147

Schermer, J. A., & MacDougall, R. (2013). A general factor of personality, social desirability, cognitive ability, and the survey of work styles in an employment selection setting. *Personality and Individual Differences, 54*, 141-144. doi: 10.1016/j.paid.2012.08.012

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits. *Journal of Cross-Cultural Psychology, 38*, 173-212. doi: http://dx.doi.org/10.1177/0022022106297299

Schütz, A. (1998). Assertive, offensive, protective, and defensive styles of self-presentation: A taxonomy. *The Journal of Psychology, 132*, 611-628. doi: http://dx.doi.org/10.1080/00223989809599293

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). New York, NY: Academic Press.

Schwartz, S. H. (1999). A theory of cultural values and some implications for work. *Applied Psychology, 48*, 23-47. doi: 10.1111/j.1464-0597.1999.tb00047.x

Schwartz, S. H. (2009a). Culture matters: National value cultures, sources, and consequences. In R. S. Wyer, C.-y. Chiu & Y.-y. Hong (Eds.), *Understanding culture:Theory, research, and Application* (pp. 127-150). New York, NY: Psychology Press.

Schwartz, S. H. (2009b). Draft users manual: Proper use of the Schwarz Value Survey. http://www.crossculturalcentre.homestead.com

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology, 26*, 92-116. doi: 10.1177/0022022195261007

Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology, 36*, 3-18. doi: 10.1111/j.2044-8309.1997.tb01115.x

Sedikides, C., Gaertner, L., & Vevea, J. L. (2005). Pancultural self-enhancement reloaded: A meta-analytic reply to Heine (2005). *Journal of Personality and Social Psychology, 89*, 539-551. doi: 10.1037/0022-3514.89.4.539

SHL. (1999). *The OPQ32 manual*. Thames Ditton, UK: SHL Group Ltd.

SHL. (2006). *The OPQ32 technical manual*. Thames Ditton, UK: SHL Group Ltd.

SHL. (2009). *The OPQ32r user manual* Thames Ditton, UK: SHL Group Ltd.

SHL. (2013). *The OPQ32r technical manual*. Thames Ditton, UK: SHL Group Ltd.

Smith, D. H. (1967). Correcting for social desirability response sets in opinion-attitude survey research. *Public Opinion Quarterly, 31*, 87-94. doi: 10.1086/267486

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology, 35*, 50-61. doi: 10.1177/0022022103260380

Smith, P. B. (2009). On finding improved ways of characterizing national cultures. In R. S. Wyer, C.-y. Chiu & Y.-y. Hong (Eds.), *Understanding culture: Theory, research, and application* (pp. 151-161). New York, NY: Psychology Press.

Smith, P. B. (2011). Communication styles as dimensions of national culture. *Journal of Cross-Cultural Psychology, 42*, 216-233. doi: 10.1177/0022022110396866

Smith, P. B., & Fischer, R. (2008). Acquiescence, extreme response bias and culture: A multilevel analysis. In F. J. R. van de Vijver, D. A. van Hemert & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 285-314). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Statistics-Netherlands. (2011). from http://statline.cbs.nl/StatWeb/publication/?DM=SLEN&PA=37325eng&D1=0-2&D2=0-4,136,151,214,231&D3=0&D4=0&D5=1,3-10,102,139,210,225&D6=14-15&LA=EN&HDR=G2,G3,G4,T&STB=G1,G5&VW=T

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of clinical psychology, 28*, 190-193.

Sturgis, P., Roberts, C., & Smith, P. (2010). *Middle alternatives revisited: How the neither/nor response acts as a face-saving way of saying I don't know*. S3RI Methodology Working Papers, (M10/01). Southampton Statistical Sciences Research Institute, Southampton.

Sutton, R. M., & Farrall, S. (2005). Gender, socially desirable responding and the fear of crime: Are women really more anxious about crime? *British Journal of Criminology, 45*, 212-224. doi: 10.1093/bjc/azh084

Tellis, G. J., & Chandrasekaran, D. (2010). Extent and impact of response biases in cross-national survey research. *International Journal of Research in Marketing, 27*, 329-341. doi: 10.1016/j.ijresmar.2010.08.003

The World Bank. (2011). World Data Bank. World Development Indicator (WDI) & Global Development Finance (GDF). Retrieved 18/02/2011, from http://www.worldbank.org/

Thunholm, P. (2001). Social desirability in personality testing of military officers. *Military Psychology, 13*, 223-234. doi: 10.1207/S15327876MP1304_3

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74*, 118-128. doi: 10.1037/0022-3514.74.1.118

Trimble, D. E. (1997). The Religious Orientation Scale: Review and meta-analysis of social desirability effects. *Educational and Psychological Measurement, 57*, 970-986. doi: 10.1177/0013164497057006007

Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: A quantitative synthesis. *American Educational Research Journal, 16*, 375-389. doi: 10.3102/00028312016004375

United Nations. (2010). UN data. Human Development Index. Retrieved 20/02/2011, 2011, from http://data.un.org/DataMartInfo.aspx#15

United Nations. (2012). UN data. Human Development Index. Retrieved 25/09/2013, 2013, from http://data.un.org/DataMartInfo.aspx#15

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*, 243-262. doi: 10.1177/1745691610369465

van de Gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries *Journal of Cross-Cultural Psychology, 43*, 1205-1228. doi: 10.1177/0022022111428083

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.

van de Vijver, F. J. R., & Meiring, D. (2011). *Social desirability among Blacks and Whites in South Africa*. Cross-Cultural Psychology Symposium. Unpublished paper. Tilburg.

van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33*, 141-156. doi: 10.1177/0022022102033002002

van de Vijver, F. J. R., & Watkins, D. (2006). Assessing similarity of meaning at the individual and country level: An investigation of a measure of independent and interdependent self *European Journal of Psychological Assessment, 22*, 69-77. doi: 10.1027/1015-5759.22.2.69

van Dijk, T. K., Datema, F., Piggen, A.-L. J. H. F., Welten, S. C. M., & van de Vijver, F. J. R. (2009). Acquiescence and extremity in cross-national surveys: Domain dependence and country-level correlates. In A. Gari & K. Mylonas (Eds.), *Quod erat demonstrandum: From Herodotus' ethnographic journeys to cross-cultural research* (pp. 149-158). Athens: Pedio Books Publishing.

van Emmerik, H., Gardner, W. L., Wendt, H., & Fischer, D. (2010). Associations of culture and personality with McClelland's motives: A cross-cultural study of managers in 24 countries. *Group & Organization Management, 35*, 329-367. doi: 10.1177/1059601110370782

van Hemert, D. A. (2003). Cross-cultural meta-analyses. *Online Readings in Psychology and Culture, Unit 2*. http://scholarworks.gvsu.edu/orpc/vol2/iss2/5

van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33*, 1229-1249. doi: http://dx.doi.org/10.1016/s0191-8869(02)00007-7

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346-360. doi: 10.1177/0022022104264126

van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*, 195-217. doi: 10.1093/ijpor/eds021

Vanhanen, T. (2007). The Polyarchy Index of Democracy. Retrieved 18/02/2011, from http://www.prio.no/CSCW/Datasets/Governance/Vanhanens-index-of-democracy/

Ventimiglia, M., & MacDonald, D. A. (2012). An examination of the factorial dimensionality of the Marlowe Crowne Social Desirability Scale. *Personality and Individual Differences, 52*, 487-491. doi: http://dx.doi.org/10.1016/j.paid.2011.11.016

Verardi, S., Dahourou, D., Ah-Kion, J., Bhowon, U., Tseung, C. N., Amoussou-Yeye, D., . . . Rossier, J. (2010). Psychometric properties of the Marlowe-Crowne Social Desirability Scale in eight African countries and Switzerland. *Journal of Cross-Cultural Psychology, 41*, 19-34. doi: 10.1177/0022022109348918

Vieluf, S., Kunter, M., & van de Vijver, F. J. R. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education, 35*, 92-103. doi: 10.1016/j.tate.2013.05.006

Wang, J. H.-Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. *Reading Research Quarterly, 39*, 162-186. doi: 10.1598/RRQ.39.2.2

Watkins, D. (1996). The influence of social desirability on learning process questionnaires: A neglected possibility? *Contemporary Educational Psychology, 21*, 80-82. doi: 10.1006/ceps.1996.0006

Watkins, D. (2000). Learning and teaching: A cross-cultural perspective. *School Leadership & Management: Formerly School Organisation, 20*, 161-173. doi: 10.1080/13632430050011407

Watkins, D. (2009). Motivation and competition in Hong Kong secondary schools: the students'perspective. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: changing contexts, changing education* (pp. 71-88). Hong Kong: Springer.

Watkins, D., & Cheung, S. (1995). Culture, gender, and response buis: An analysis of responses to the Self-Description Questionnaire. *Journal of Cross-Cultural Psychology, 26*, 490-504. doi: 10.1177/0022022195265003

Watkins, D., McInerney, D., Akande, A., & Lee, C. (2003). An investigation of ethnic differences in the motivation and strategies for learning of students in desegregated South African schools. *Journal of Cross-Cultural Psychology, 34*, 189-194. doi: 10.1177/0022022102250563

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070. doi: 10.1037/0022-3514.54.6.1063

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods, 15*, 96-110. doi: 10.1037/a0018721

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34*, 702-722. doi: 10.1177/0022022103257070

Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *Journal of Experimental Education, 72*, 331-346.

Xu, Y., Zhang, L., & Hee, P. (2014). Parenting practices and shyness in Chinese children. In H. Selin (Ed.), *Parenting Across Cultures* (Vol. 7, pp. 13-24): Springer Netherlands.

Yu, J., & Murphy, K. R. (1993). Modesty bias in seld-ratings of performance: A test of the cultural relativity hypothesis. *Personnel Psychology, 46*, 357-363.

Summary

Response styles have been studied since the 1950s, yet its psychological meaning and implications on validity of data are still under debate. Challenges in research on response styles include the different operationalizations of specific response styles, lack of validity measures that are less susceptible to response styles, and the inconsistency in findings of their correction effects. This dissertation aims to advance our understanding of response styles from a cross-cultural perspective by (1) integrating different response styles to a general factor, (2) establishing the nomological network of response styles with various validity measures at both individual and cultural level, and (3) exploring the implications of response style effects in cross-cultural surveys. Nine empirical studies were carried out, covering three themes.

The first three chapters address the integration of specific response styles among ethnic groups in the Netherlands. Chapter 2 aims to confirm a General Response Style. Specifically, we addressed commonalities and differences of acquiescence, extremity, midpoint responding, and socially desirable responding that can be taken to constitute a single underlying response style. Participants were 548 Dutch nationals and 1,116 first- and second-generation immigrants of Western and Non-Western origins in the Netherlands. Self-report measures of the four response styles, and personality traits were administered. Conventional, indirect measures of acquiescence, extremity, and midpoint responding were also calculated. A multigroup confirmatory factor analysis showed support for a General Response Style factor with positive loadings of extremity and socially desirable responding, and negative loadings of acquiescence and midpoint responding. The response style factor was strongly associated with personality (notably the "Big One" factor). These findings support a view that there is a General Response Style factor and that, in addition, each response style has some unique meaning. The ethnic groups differed significantly on response style use, with Non-Western immigrants showing higher acquiescence and midpoint responding than the other groups.

Chapter 3 deals with the replication of the General Response Style factor and a test of the domain specificity of response styles. A total of 5,457 Dutch residents (mainstreamers and Western and non-Western immigrants) were sampled. Self-report measures of the four response styles, values, personality, self-regulation, cognition, positive life outcomes, and political views were administered. Conventional, indirect measures of acquiescent, extreme, and midpoint responding were calculated. The General Response Style factor was successfully replicated. This factor was found to be most strongly associated with personality

and cognition, and least with political views. However, corrections for response styles were not consequential in that correlations among psychological variables were not affected by such corrections and in that the size of gender, age, and ethnic differences were not affected by score corrections.

Chapter 4 further extends the General Response Styles to a self-presentation style in a longitudinal study. We investigated how response styles, personality traits, and values can be taken as manifestations of self-presentation styles in self-reports, and how self-presentation affects other self-report measures over time. Data on values and character traits at three time points across five years collected among a national representative sample in the Netherlands were utilized. A General Response Style factor consisting of extreme, socially desirable, and midpoint responding, a general factor of personality from the International Personality Inventory, and a general value factor from the Rokeach Value Survey were extracted, all of which showed scalar invariance across time. A latent self-presentation factor defined by the three general factors at each time point, and its stability and changes across time points was modeled. All the three general factors loaded positively on the self-presentation factor. The latent mean of the self-presentation factor became smaller over time, yet effects of its impact on the relationships among various psychological variables remained small and stable over time.

The next three chapters address cross-cultural variations in response styles. Chapter 5 focuses on social desirability. We investigated the dimensionality, measurement invariance, and cross-cultural variations of social desirability, using an adapted version of the Marlowe-Crowne scale among 3,471 university students from 20 countries. A two-dimensional structure was revealed in the pooled sample, distinguishing enhancement (endorsement of positive self-description) and denial (rejection of negative self-description). The factor structure was supported in most countries; medium-sized item bias was found in two denial items. In a multilevel analysis, we found that (1) there was more cross-cultural variation in denial than enhancement; (2) females tended to score higher on enhancement whereas males tended to score higher on denial; (3) the Human Development Index, an indicator of socioeconomic development, was the best (negative) predictor of denial, and (4) both enhancement and denial seemed to be associated with country-level values and personality pertinent to “fitting in”.

Chapter 6 extends the individual-level General Response Style to country level. Individual- and country-level indexes of acquiescent, extreme, and midpoint responding were constructed from eight multicountry surveys for correlation and multilevel analysis. At both

levels, we confirmed the General Response Style factor with a positive loading of extreme response style, a negative loading of midpoint response style, and with acquiescent response style in between. At country level, the General Response Style was negatively related to socioeconomic development and the percentage of atheists, and positively related to aggregated values and personality traits pertinent to "fitting in" and avoidance of ambiguity. At individual level, the General Response Style was positively associated with age and negatively associated with education.

Chapter 7 examines the shared and unique meaning of acquiescent, extreme, midpoint, and socially desirable responding in association with the OPQ32, a forced-choice format personality measure designed to be less affected by these response styles, compared to personality inventories with Likert scales. Country-level response style indexes were derived from six waves of the International Social Survey Programme and from a meta-analysis of a social desirability scale. In the country-level correlational analysis the four response styles formed a General Response Style factor which was positively associated with (1) dominance (vs. submission) in interpersonal relationships, (2) competitive (vs. modest and democratic) feelings and emotions, and (3) data rational thinking. In a multilevel analysis, age showed a positive and education a negative effect on the individual-level General Response Style. Negative effects of country-level socioeconomic development and individualism and positive effects of competitiveness and data rational thinking on the individual-level response style were found.

The last three empirical chapters address the implications of response styles on cross-cultural score differences. Chapter 8 zooms in on the effects of acquiescence and social desirability in association with the Schwartz Value Survey and other international datasets. We found that affluence explained a substantial proportion of the variance in the association of response styles with value scores in various surveys. Value score standardization in the Schwartz Value Survey had some effect on the correlations of acquiescence with various value types, but only limited effects on social desirability.

In Chapter 9 we investigate the integration of response styles and their effects on self-reports among 76,887 teachers in 18 countries in the Teaching and Learning International Survey (TALIS). Socially desirability (with a positive and a negative impression management factor) and 17 core constructs related to the teaching profession were measured with Likert scales; extreme and midpoint response styles were constructed. A General Response Style was extracted with socially desirable and extreme response styles as positive indicators and midpoint response style as a negative indicator. This General Response Style was more

strongly correlated with constructs of personal involvement, such as teacher efficacy and job satisfaction, than constructs with less personal involvement, both at individual and country level; however, its correction had negligible effects on the size of cross-cultural differences and the country rank order in any construct.

Chapter 10 compares the effects of response styles and other methods to correct for scale usage in a recurrent paradox in the data of the Programme for International Student Assessment (PISA): within each participating country, there is a positive correlation between students' learning motivation and achievement; when aggregating the data at country level, this correlation becomes negative. Using PISA data across 64 countries, we investigate the association of motivation and achievement within and between countries, and attempt to explain the paradox with three measures indicative of culturally preferred scale usage: cultural response style, overclaiming, and anchoring vignettes. We confirmed the paradox and found that the three measures, in particular anchoring vignettes, could partially explain the negative association between motivation and achievement at country level.

The central questions that this dissertation tries to answers are: How can response styles be reliably and validly measured? Are they styles, or are they substance? And how should we deal with response styles in survey data? The main findings are as follows: (1) There is a General Response Style that can integrate specific response styles. With extremity and social desirably as positive indicators and acquiescence and midpoint responding as negative indicators at both individual and culture level, this General Response Style represents the tendency of response amplification versus moderation. Replicable across multiple measures and samples, this General Response Style is believed to help create consistency in findings of response styles. (2) The General Response Style is meaningfully related to individual and cultural characteristics. At individual level, the General Response Style is strongly related to positive personality traits and positive life outcomes such as self-esteem and life satisfaction; moderately related to horizontal and vertical collectivism, horizontal individualism, and emotion appraisal; and weakly associated with political views. It points to the domain specificity of response styles: the higher the personal relevance, the stronger the General Response Style. At cultural level, this General Response Style is negatively associated with socioeconomic development of a country and positively related to values pertaining to "fitting in" and ambiguity avoidance, and country personality traits such as dominance, competitiveness, and data rationality. The nomological network of specific response styles could be inferred based on their loadings on this General Response Style. (3) Effects of response style correction in cross-cultural contexts are mild. Correcting for self-

presentation (with the General Response Style as an indicator) has weak effects on the intercorrelations among various psychological variables, and the correction of the General Response Style hardly resulted in dramatic changes of country rankings, or the effect sizes of cross-cultural differences. All in all, caution is needed in response style correction.

Acknowledgements

My PhD journey has been both challenging and rewarding, and I am deeply thankful to those who have guided and accompanied me through the last four years.

Fons, it has been a great privilege to work with you. Thank you for all your support, guidance, and patience. You are the best supervisor that I can ever ask for. Being your student for four years, I aspire one day to be a supervisor like you for my students.

Alejandra, I truly appreciate your belief in me, your kind help and valuable feedback. I remember how unconfident I was at the beginning of my PhD, and you were always there to encourage me and help me out. I got so much inspiration from you, and I hope to work with you again in the future.

Ype, I am grateful for your insights and suggestions for my project, and all the stimulating conversations.

My gratitude also goes to my colleagues and friends in Tilburg. Atha and Michael, thank you for your valuable feedback in our research meetings. Amina, your academic commitment and hard-working spirit is inspiring. Elif, Byron, Maja, and Snezana, you are great officemates and friends. Ozgur, thank you for all the talks and walks in the forest. Arzu, Isabel, and Radosveta, thank you for your friendship and support in my project. As an immigrant in this country, I feel as home because of our close-knit cross-cultural group.

I want to express appreciation to the participants who took part in my studies, to my coauthors in the manuscripts, and to the members of the CentERdata for their support with data collection.

I have wonderful support from my family and friends thought out the project. My parents have always been supportive in my academic career choice and believed in me. Greta, my mother-in-law, thank you for caring for me. Bez and Roy, thank you for supporting my decision to leave Australia and pursue my PhD in the Netherlands.

Bernd, my dear husband, it is destine that we met the first week I arrived in Tilburg and started a life together. My research is about creating consistency in response styles, and you are the one who helps me create consistency in life. You have always been considerate and supportive, for which I am truly grateful. This thesis is dedicated to you.

Additional Publications

He, J., & van de Vijver, F., J. R., (accepted). Bridging emic and etic approaches in cross-cultural management research. Chapter in *Routledge Companion to Cross-Cultural Management*.

He, J., & van de Vijver, F., J. R., (accepted). Cross-cultural methods of research. In H. L. Miller (Ed.). *The SAGE encyclopedia of theory in psychology*. Thousand Oaks, CA: SAGE.

He, J., & van de Vijver, F. J. R. (2014). The value of keeping an open eye for methodological issues in research on resilience and culture. In L. C. Theron, L. Liebenberg & M. Ungar (Eds.), *Youth resilience and culture: Commonalities and complexities*. (pp. 189-201). Dordrecht: Springer.

He, J., & van de Vijver, F. J.R. (2013). Methodological issues in cross-cultural studies in educational psychology. In G. A. D. Liem & A. B. I. Bernardo (Eds.), *Advancing cross-cultural perspectives on educational psychology: A festschrift for Dennis McInerney* (pp. 39-56). Charlotte, NC: Information Age Publishing.

He, J., & van de Vijver, F. J. R. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, Unit 2*. Retrieved from http://scholarworks.gvsu.edu/orpc/vol2/iss2/8